LICENTIATE THESIS

Towards Quality of Experience for Industrial Internet of Things

Dimitar Minovski

Pervasive and Mobile Computing

Luleå University of Technology

Towards Quality of Experience for Industrial Internet of Things

Dimitar Minovski

Dept. of Computer Science and Electrical Engineering
Luleå University of Technology
Luleå, Sweden

Supervisors:

Chriter Åhlund, Karan Mitra, Irina Cotanis

Printed by Luleå University of Technology, Graphic Production 2020

ISSN 1402-1757
ISBN 978-91-7790-558-5 (print)
ISBN 978-91-7790-559-2 (pdf)

Luleå 2020

www.ltu.se

To my family

Abstract

Today's research on Quality of Experience (QoE) mainly addresses multimedia services, where the end-users' subjective perception is the prime factor of determining the QoE. With the introduction of the Internet of Things (IoT), there is a need for new ways of evaluating the QoE. Emerging IoT services, such as remotely-controlled operations, autonomous vehicles (AVs), and energy management, are more complex, creating additional quality requirements emerging from the machine-to-machine (M2M) communication and autonomous processes. One challenge, as an extension of the legacy QoE concept, is understanding the perception of end-users QoE in the context of IoT services. For instance, within the current state of the art in QoE it is not clear how intelligent machines can impact end-users' QoE, but also how end-users can alter or affect an intelligent machine. Another challenge is the quality evaluation of the M2M and systems that enable the machines to run by themselves. Consider a self-driving vehicle, where multiple autonomous decisions are simultaneously made as a result of predictive models that reason on the vehicle's generated data. An evaluation of the predictive models is inevitable due to abundance of the potential sources of failures. A quality degradation of the IoT hardware, the software enabling autonomous decision, and the M2M communication can raise life-threatening concerns, directly impacting the end-users' QoE.

In this thesis, we argue for a paradigm shift in the QoE area that understands the relationships between humans and intelligent machines, as well as within the machines. Our contributions are as follows: first, we introduce the term Quality of IoT-experience (QoIoT) that extends the conventional QoE approaches in covering IoT services. Within QoIoT, we consider a quality evaluation from the perspective of the end-users, as well as from intelligent machines. The end-user's perception is captured by following the conventional QoE approaches, while regarding intelligent machine we propose the usage of objective metrics to describe their experiences and performance. As our second contribution, we propose a novel QoIoT architecture that consists of a layered methodology in order to determine the overall QoIoT. The QoIoT architecture, firstly, models the data-sources of an IoT service, classified within four layers: physical, network, application, and virtual. Secondly, the architecture proposes three layers for measuring the QoIoT by considering Quality of Data (QoD), Quality of Network (QoN), and Quality of Context (QoC), with QoC being the prime layer in measuring the objective performance metrics. Finally, the third contribution of this thesis considers a case-study of cellular IoT, involving autonomous mining vehicles, which we utilize to achieve a preliminary results that validate the proposed QoIoT architecture.

Contents

Acknowledgments

I would like to thank my advisors. My sincere gratitude goes to Professor Christer Åhlund for giving me the opportunity to become a PhD student and all the lessons I learned along my journey, both in professional and private life. I would like to thank my co-advisor, Associate Professor Karan Mitra for the collaboration and support. Many thanks to my fellow PhD students and rest of the staff at the university.

This thesis would not have been possible without the support from our industrial partner, InfoVista Sweden AB. I had the pleasure to work with amazing colleagues. I am extremely gratefully for our collaboration. Special thanks goes to Ulf Marklund, Irina Cotanis, Niclas Ögren, and Per Johansson.

Finally, and most importantly, I would like to thank my family, my girlfriend Mimmi and her family, and my friends for everything they have done for me, their support, inspiration, and sacrifices through life. I love them and proud to have such incredible people in my life.

Skellefteå, March 2020
Dimitar Minovski

Part I: Thesis Introduction

Chapter 1

Introduction

1.1 Introduction

Internet of Things is defined as a global infrastructure for the information society, enabling services by interconnecting physical and virtual objects based on existing and evolving information and communication technologies [41]. The ever-growing demand for connected devices went from around 500 million in 2003, to 8.6 billion in 2018 and forecasted 22.3 billion by 2024, excluding mobile phones [25]. The applications for IoT has grown from simple sensing devices into enabling advanced services for remote controlling, management, and maintenance of physical and virtual processes [33]. Few examples of the wide range of emerging IoT applications are shown on Figure 1.1, including digitizing society (e.g. smart city, health-care), automated industrial production, and intelligent transportation. The rapid increase in deployed devices comes with the costs of series of challenges, including network scalability, data management, power consumption, and security [61]. For instance, there are demands on the access networks for developing new IoT-tailored communication protocols [8]; and demands on IoT platforms for bringing the computation and virtual storage closer to the devices [71].

As IoT is becoming more pervasive [46], it expands to services with a wide range of unique properties and quality requirements. Consider Autonomous Vehicles (AVs) as an example, where the focus is on fast and accurate processing of sensory data to avoid collisions [15]; while IoT-enabled remotely controlled vehicles impose stringent latency requirements on the machine-to-machine (M2M) communications [74]. Such emerging IoT services are engaging closely to human life-style and privacy, where failures in the M2M type communications can pose life-threatening risks, as well as risks of causing business losses. The current practice of dealing with fulfillment of network quality requirements during run-time of a service is achieved through evaluation of Quality of Service (QoS) metrics [34]. For instance, the QoS metrics can describe the network delay, jitter, packet drops, and bandwidth. However, the QoS metrics narrowly capture the end users' perception of the service and may not reflect their overall service accept-

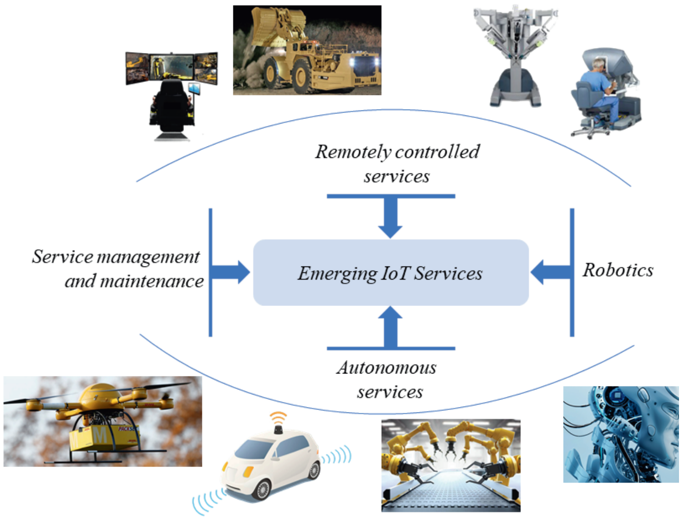


Figure 1.1: The emerging IoT use-cases.

ability and experience [59]. To highlight the end user's point-of-view, the concept of Quality of Experience (QoE) was defined as: *"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 the light of the user's personality and current state."* [16]. Since then, the research on QoE has focused from mapping QoS metrics to users' QoE [27]; to understanding and defining factors influencing QoE [59]; and to building QoE management models [60].

The research on QoE management is addressed from two distinct, but often complementary perspectives, application and network [73]. The network QoE management is mainly an extension of the legacy Quality of Service (QoS) techniques, measuring metrics to describe network performance and the way they impact the end-user perception [27]. Whereas application QoE considers social, cognitive, and contextual metrics to produce a subjective score that ranks the end-users' experience [59]. However, the conventional QoE management mainly targets multimedia services, such as video, voice, gaming, and web applications [73]. Evaluating QoE in IoT services, especially in industrial domain, deviates from the conventional paradigm, as the scope goes beyond measuring single end-user's perception of a multimedia. Consider a simple scenario of regulating the temperature in a building. Herein, the network QoE would assess the service in terms of sending timely updates to a end-user in cases of remote supervision [73]. Whereas the application QoE can assess the end-user's subjective perception of the service in terms of feeling pleasantness in the environment [73]. The end-user's experience in this case depends on measurements from multiple interconnected sensors, managed from a centralized unit. Therein, evaluating the precision, accuracy, and quality of the sensory data is currently lacking in the QoE management models [80]. Inability to detect hardware and software failures in the IoT objects, and thus their implications on QoE, backfires when evaluating performance and QoE of more advanced services. Emerging IoT services include M2M type communication among physical and virtual objects, enabling intelligent machines of autonomous decision-making based on interconnected predictive models [77]. Having a poor understanding of how an autonomous service can impact the end-users QoE will lead to reduction of their situation awareness and potentially have catastrophic repercussions [24], especially in AVs case. Thus, in this thesis, we define and model the end-to-end quality of experience in an IoT service. In particular, we identify the main components and interactions within an autonomous service and propose metrics to evaluate the experience from two perspectives: *(1) Human*, by extracting the end-users' QoE; and *(2) Machines*, by translating the overall machine performance to business oriented metrics.

1.2 Research Motivation

Human QoE. The research on QoE focuses on quantitative representation of subjective human perception of the service [60]. For instance, an end-user watches a video stream via the internet, downloading frames from a remote video server, and then rates its subjective perception of the service. Therein, the main factors affecting the QoE are the network performance (e.g. QoS), subjective (e.g. mood, enjoyment, emotions) and objective human factors (e.g. Human visual system (HVS)). The state of the art in QoE mainly evaluates a end-user perception of multimedia services, such as video, voice, gaming, and web [60]. However, in the emerging IoT services the human-computer interaction is on different dimension in terms of how the new technology will impact the end-users [77]. Consider an AV, where decisions by the software can impact multiple end-users' QoE simultaneously, ranging from the driver, passengers, pedestrians, and people in surrounding vehicles. Moreover, multiple end-users can affect and alter the real-time decisions by the AV. For instance, a driver can override the AV, a passenger can re-adjust the AV's settings, or a pedestrian crossing a street can stop the AV. Thus, the relationships between the end-users and the autonomous service, such as an AV, are much more complex [77]; The conventional approaches of measuring QoE and its metrics require an extension in covering perception of multiple QoEs depending on the contextual situations.

Service QoE. A fundamental characteristic of an IoT service is the coalescence with functionalities provided by other services [46], leading to complex service chains, known as System-of-Systems (SoS) [14], with possibly even hundreds of interconnected services offered by different third parties, each with their own business incentives. Consider a mining company with a fleet of AVs digging and excavating natural resources (e.g. ore). Now consider an emergency situation in the mine, such as fire, explosion, or flood. The risk assessment comprises of an alarm service that ensures the safety of the miners, AVs, and other equipment. The end-to-end alarm service consists of multiple sensors to detect smoke, humidity, and temperature, as well as of autonomous processes for dealing with the emergency. For instance, an alarm is triggered in case of a fire in the mine, which is followed by a series of chained events, such as reasoning on positioning data to locate the miners and AVs and compute their most optimal rescue route. Then,

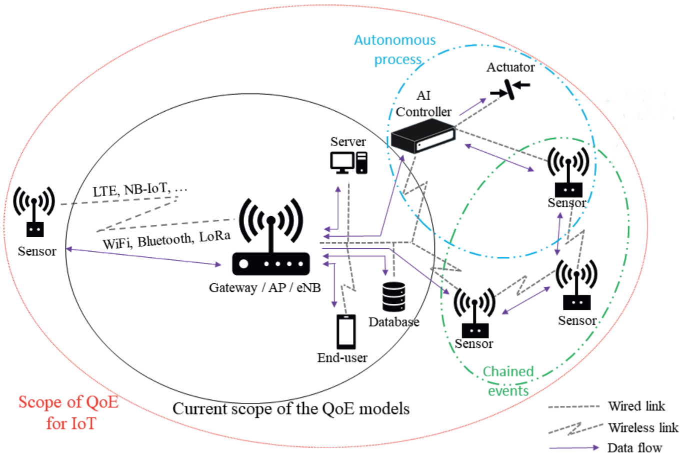


Figure 1.2: QoE in an end-to-end IoT service.

the service must actuate by notifying the miners and AVs for their individual routes in a timely manner and contact additional external services to deal with the emergency situations (e.g. fire department). The safety of the miners and machines in this case rely on complicated dependencies on the data-sources and M2M communication. Consider Figure 1.2 as an example of such end-to-end IoT service, consisting of service chains and autonomous processes. Therein, an output from one sensor is an input to another sensor, followed by an autonomous reasoning and actuating, creating complex chains with strong dependencies on the accuracy and quality of the produced data by each sensor and process. In current practice, the quality and performance of such complex SoS is controlled on an ad-hoc basis, while the consequences of failures are not well understood [77]. Therefore, there is a need for extension of conventional QoE techniques to cover complex SoS. In particular, there is a need of automated techniques for measuring the impact of chained services on the business-oriented metrics and QoE, understanding the implication of failures within and offering root-cause analysis.

Machine QoE. AI is seen as a key enabler for autonomous services [46], which is highlighted in the Figure 1.2. Going into the essence of AI, it is defined as *"the science of making machines capable of performing tasks that would require intelligence if done by humans"* [58], while recent studies refer to AI as *"similar or the same kind as human intelligence exhibited machines"* [50]. The research on AI goes as far as defining machine emotions [65] and human-robot interaction [31]. Thus, one may envision that AI-enabled machines may, just like humans, perceive *experience*, which arguably is different compared the humans (e.g. QoE) [35]. AI typically runs as a software controller (Figure 1.2), controlling individual processes or even a complete end-to-end service, making decisions based on internal and external data-sources [50]. In addition, a single decision by the AI can be based on the M2M communication among multiple predictive models, creating additional dependencies on the accuracy of the outputs as a result of a prediction. For instance, consider a predictive AV model for overtaking a vehicle on the road which reasons on the produced predictions from multiple individual models, such as for detecting moving objects, line-keeping, collision avoidance, and speed control [9, 26]. The M2M interaction among those predictive models is itself a machine experience in a sense of how each AI process perceives the data inputs. A faulty prediction by any of the models may occur as a results of a faulty input data, as well as a result of a inaccurate, biased, and overfitted or underfitted AI model, degrading the overall service performance regarding self-driving [77]. In that direction, the critical challenge remains in identifying the main data-sources (e.g. raw data and data as outcome of the prediction models) and then produce metrics for benchmarking the complete AI system. The benefits can be three-fold: (1) Improving the end-users' QoE and awareness of the machine's state, enabling the end-user to override the AI's decisions in the right time or tweak its settings; (2) Facilitate the AI's self-learning by using the benchmarking metrics for labeling the raw generated data; (3) Ability to trigger root-cause analysis on a measured poor performance and identify the source issues.

1.3 Research Questions

This section presents the selected research questions based on the state of practice and state of the art review. In the following, we describe two research questions which this thesis attempts to answer:

RQ1: How to define Quality of Experience for an IoT service?

QoE evaluation models have been traditionally developed for multimedia services, centered around the end-user's perception. However, the emerging IoT services bring additional requirements which infer that the constituted QoE concept require an extension. First, the QoE models should consider the quality of the produced data by the IoT objects, as the QoE can be degraded by a single faulty measurements by an object. Second, an IoT service may include M2M communication among multiple objects, based on which an intelligent machine can gather and process the data to make autonomous decisions. Therein, a QoE model should understand and evaluate the impact of M2M communication on the overall service performance. Third, the relationship between the end-users and IoT service is on different dimension regarding the metrics that can impact the QoE. For instance, a mission-critical IoT service, such as in health-care, need to evaluate the impact of intelligent machines regarding safety and reliability metrics. Therefore, with **RQ1** we raise the need of extending the working QoE definition, proposed in [16], to cover IoT services.

RQ2: How to develop and validate an architecture for Quality of Experience in an IoT service?

The existing QoE models, surveyed by Skorin et al. [73], consist of architectures for evaluating QoE that typically process application and network layer data to produce QoE metrics as an output. In that direction, the need to extend the QoE definition, as addressed with RQ1, would also require an extension of the state of the art architectures.

Therefore, we raise the need of an architecture that models the IoT requirements from RQ1 and quantify the experiences perceived by the end-users and intelligent machines.

1.4 Research Methodology

This section presents an overview of the research methodology that is used in this thesis, illustrated on Figure 1.3. According to Kothari [49], research methodology is a method(s) to analytically explain and solve a research problem. Our research process follows the three steps research framework proposed by Holtz et al [36], although slightly modified to serve our research context of implementing a real-world solutions.

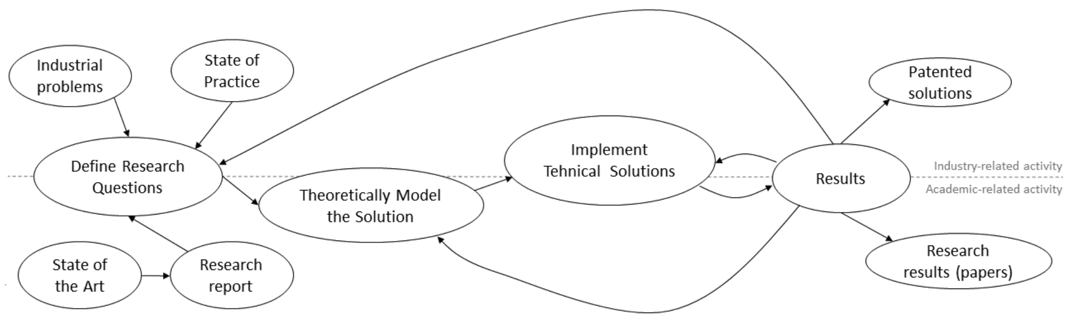


Figure 1.3: The research process in this thesis.

Our research process, shown on Figure 1.3, is a results of a close cooperation with our industrial partner InfoVista AB, whose expertise provides valuable contributions in every step of the process. The first step of our research process is to define research questions. The initial point of discussion was set on two wide areas of research - IoT and cellular networks, to match the state of practise within our industrial partner. Then, a specific set of topics and technologies were selected for further research based on: (1) Industrial problems that need to be addressed, according to the state of practise within our industrial partner; and (2) Their disruptive impact within the two initially defined areas, according to the state of the art literature. As a result, a research report was produced that addresses the state of the art literature describing key challenges in the following topics and technologies: (1) IoT (e.g. architectures); (2) Access network technologies within IoT (e.g. LoRa, Narrowband-IoT); (3) Future network technologies within IoT (Software Defined Networks, edge computing, caching); and (4) QoE (e.g. QoE management and provisioning). Based on the state of the art report, a series of discussions followed in cooperation with the industrial partner to select the most valuable challenges and form research questions as a basis for this thesis, discussed in Section 1.3.

During the second step of our research process (Figure 1.3) we attempted to answer the RQ1, from Section 1.3, which required thorough review of the state of the art in the area of QoE for IoT. The review includes an understanding of the current and upcoming IoT services, to which the state of the art QoE principles should be applied. We then attempted to refine the existing theories that define QoE and build a conceptual architecture as a solution to the defined research questions. Thus, the process of developing the architecture consisted of the following research methods: *(1) formulative;* and *(2) descriptive research* [36, 49]. In addition, we utilize a *case-study research method* [49] of autonomous vehicles to identify input and output metrics of the proposed conceptual model for evaluating QoE for IoT services. The process of developing the conceptual model is not linear and requires iterative refinements based on the technical implementations and validation of the achieved results.

During the last step of the research process, we considered the *quantitative research method* [36] to collect a real-life data within the case-study by implementing technical solutions from which our research and patented results emerge. The technical solutions are developed on a pre-selected parts of the proposed conceptual architecture and serve as a preliminary analysis to validate the model in industrial settings.

1.5 Thesis Contribution

In this section we present the main contributions from this thesis. Figure 1.4 presents a high-level overview of the contributions in terms of the way research results address the stated research questions from the Section 1.3. The main contributions include: (1) Extending the QoE definition in context of IoT services; (2) Developing an architecture that models the quality of IoT experience; and (3) Conducting a preliminary analysis for validating the proposed architecture.

In the following, we list the produced research results and then in detail discuss the individual contributions.

- **Paper A.** *Quality of Experience for the Internet of Things.* **Dimitar Minovski**, Chrsiter Åhlund, Karan Mitra, Roman Zhohov. IEEE IT Professional Magazine. Accepted on January 6, 2020.

  - **Abstract.** The Internet of Things (IoT) brings a set of unique and complex challenges to the field of Quality of Experience (QoE) evaluation. The state-of-the-art research in QoE mainly targets multimedia services, such as voice, video, and the Web, to determine quality perceived by end-users. Therein, main evaluation metrics involve subjective and objective human factors and network quality factors. Emerging IoT may also include intelligent machines within services such as health-care, logistics, and manufacturing. The integration of new technologies such as Machine-to-Machine communications and artificial intelligence within IoT services may lead to service quality degradation caused by machines. In this article, we argue that evaluating QoE in the IoT services should also involve novel metrics for measuring the performance of the machines alongside metrics for end-users' QoE. This article extends the legacy QoE definition in the area of IoT and defines conceptual metrics for evaluating QoE using an industrial IoT case-study.

	RQ1	RQ2
Paper A	✓	
Paper B	✓	✓
Paper C		✓
Paper D		✓

Figure 1.4: Mapping between the included papers and the research questions.

- **Authors' individual contribution.** I was the main driver of the paper. I reviewed and analyzed the state of the art on QoE. I proposed the definitions regarding QoME and QoIoT, with which we extend the current definition of QoE to cover IoT services. I wrote the complete paper. Christer Åhlund and Karan Mitra, as my main supervisors, were taking part in the complete process of building the proposed definitions, contributing with constructive feedback and discussions that led to the current form of the definition. Also, they both helped with formulating the main message of the paper and commented the paper's text to improve its structure and presentation. Roman Zhohov was involved during the initial discussions regarding the paper's idea and scope.

- **Paper B.** *Modeling Quality of IoT Experience in Autonomous Vehicles.* **Dimitar Minovski**, Chrsiter Åhlund, Karan Mitra. IEEE Internet of Things Journal. Accepted on February 10, 2020.

  - **Abstract.** Today's research on Quality of Experience (QoE) mainly addresses multimedia services. With the introduction of the Internet of Things (IoT), there is a need for new ways of evaluating the QoE. Emerging IoT services such as autonomous vehicles (AV) are more complex and involve additional quality requirements, such as those related to machine-to-machine communication that enables self-driving. In fully autonomous cases, it is the intelligent machines operating the vehicles. Thus, it is not clear how intelligent machines will impact end-users QoE, but also how end-users can alter and affect a self-driving vehicle. This article argues for a paradigm shift in the QoE area to cover the relationship between humans and intelligent machines. We introduce the term Quality of IoT-experience (QoIoT) within the context of AV, where the quality evaluation, besides end-users', considers quantifying the perspectives of intelligent machines with objective metrics. Hence, we propose a novel architecture that considers Quality of Data (QoD), Quality of Network (QoN), and Quality of Context (QoC) to determine the overall QoIoT in the context of AVs. Finally, we present a case study to illustrate the use of QoIoT.
  - **Authors' individual contribution.** I was the main driver of the paper and wrote the complete paper. I designed the proposed architecture and mapped it to the case-study of autonomous vehicles. Christer Åhlund and Karan Mitra

were involved in the complete process of developing the proposed architecture, supervising the work, and contributing with constructive feedback and discussions that led to the current version of the architecture.

- **Paper C.** *Real-time Performance Evaluation of LTE for IIoT.* Roman Zhohov, **Dimitar Minovski**, Per Johansson, Karl Andersson. In 2018 IEEE 43rd Conference on Local Computer Networks (LCN) (pp. 623-631).
  - **Abstract.** Industrial Internet of Things (IIoT) is claimed to be a global booster technology for economic development. IIoT brings bulky use-cases with a simple goal of enabling automation, autonomation or just plain digitization of industrial processes. The abundance of interconnected IoT and CPS generate additional burden on the telecommunication networks, imposing number of challenges to satisfy the key performance requirements. In particular, the QoS metrics related to a real-time data exchange for critical machine-to-machine type communication. This paper analyzes real-world example of IIoT from a QoS perspective, such as remotely operated underground mining vehicle. As part of the performance evaluation, a software tool is developed for estimating the absolute, one-way delay in end-to-end transmissions. The measured metric is passed to a machine learning model for one-way delay prediction based on LTE RAN and radio measurements using commercially available cutting-edge software tool. The achieved results prove the possibility to predict the delay figures using machine learning model with coefficient of determination up to 90%.
  - **Authors' individual contribution.** Roman Zhohov and I were the main drivers and contributors to the paper. Roman was at the time a master's student that worked under my and Karl Andersson's supervision. I initially designed the study with its main goals. Roman developed the software tool for the experiments and we jointly performed the experiments. Roman also developed the machine learning model, while we jointly performed the data analysis. I wrote three complete sections in the paper. Further on, I extended the work of this study by submitting a patent solution to the European patent office, being the main driver and writing the complete patent. Per Johansson was supervising the technical part of the study, providing feedback on the machine learning modeling.

- **Paper D.** *Analysis and Estimation of Video QoE in Wireless Cellular Networks using Machine Learning.* **Dimitar Minovski**, Chrsiter Åhlund, Karan Mitra, Per Johansson. In 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX) (pp. 1-6).
  - **Abstract.** The use of video streaming services are increasing in the cellular networks, inferring a need to monitor video quality to meet users' Quality of Experience (QoE). The so-called no-reference (NR) models for estimating

video quality metrics mainly rely on packet-header and bitstream information. However, there are situations where the availability of such information is limited due to tighten security and encryption, which necessitates exploration of alternative parameters for conducting video QoE assessment. In this study we collect real-live in-smartphone measurements describing the radio link of the LTE connection while streaming reference videos in uplink. The radio measurements include metrics such as RSSI, RSRP, RSRQ, and CINR. We then use these radio metrics to train a Random Forrest machine learning model against calculated video quality metrics from the reference videos. The aim is to estimate the Mean Opinion Score (MOS), PSNR, Frame delay, Frame skips, and Blurriness. Our result show 94% classification accuracy, and 85% model accuracy (R^2 value) when predicting the MOS using regression. Correspondingly, we achieve 89%, 84%, 85%, and 82% classification accuracy when predicting PSNR, Frame delay, Frame Skips, and Blurriness respectively. Further, we achieve 81%, 77%, 79%, and 75% model accuracy (R^2 value) regarding the same parameters using regression.

– **Authors' individual contribution.** I was the main driver of the paper. I designed the experimental setup, developed the software tool for the experiment, and conducted the experiments. Also, I pre-processed and analysed the data from a machine learning point of view, in developing the model for estimating the video streaming quality. I wrote the complete paper. Christer Åhlund and Karan Mitra were involved during the experimental design and data analysis part, providing valuable feedback. They also took part of the writing process of the paper, with feedback on the paper's structure and presentation. Per Johansson was supervising the technical part of the study, providing feedback on the data analysis and machine learning modeling.

1.5.1 *Contribution 1:* Defining Quality of IoT experience

The Paper A [57] attempts to answer the RQ1 by first reviewing the state of the art in QoE area and identifying the main challenges of assessing QoE for IoT. We identify intelligent machines and M2M communications as the key areas that need to be included in the novel, extended, definition of QoE for IoT. The M2M communication enables IoT objects to be orchestrated by an intelligent machine (e.g. AI controller on Figure 1.2), which can make automated decision based on the gathered data and multiple predictive models [30]. For instance, an AI-powered controller can autonomously change a room temperature or navigate a self-driving vehicle. Herein, the challenges come with the ability to predict outcomes based on the data-sources, a process that is prone to errors due to: (1) Network issues affecting the M2M communication; (2) Hardware and software bugs and failures; (3) Poor accuracy of the prediction models. Thus, as part of the first contribution, in Paper A [57], we acknowledge the need of evaluating the performance of autonomous processes and M2M communication, by proposing a definition for Quality of Machine Experience (QoME), which states as follows:

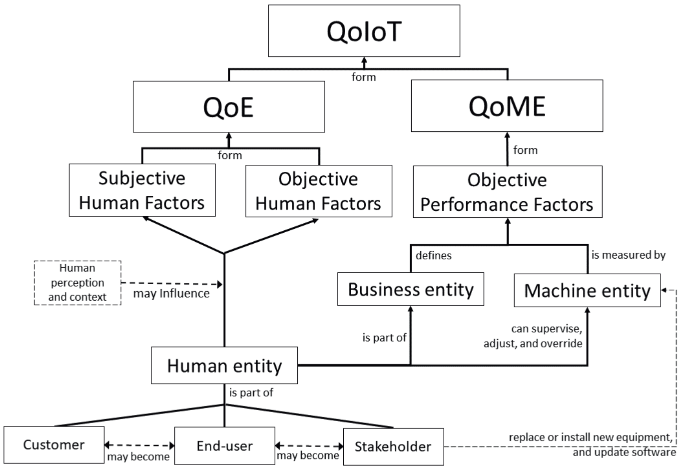


Figure 1.5: Defining QoIoT.

Definition 1. (QoME) *Objective metrics that measure the quality and performance of the machine's processes and its decisions.*

The formation of the QoME metric is service specific and should come out from the service requirements. Paper A [57] describes a industrial case-study with remotely controlled mining vehicles, with abilities to also execute tasks autonomously, based on which QoME metrics are proposed. For instance, benchmarking the progress of a machine processes (e.g. loading the ore) towards a particular business objective, such as productivity, safety, and efficiency is one form of QoME metrics. However, end-users take active participation in the IoT service and interact with the intelligent machines, such as supervising, adjusting, and controlling the vehicle from distance. Adding a layer of intelligence for autonomous decision-making in IoT services complicates the QoE evaluation process, since such decisions may also impact end-user's perspective. In such cases, the end-users' QoE can be affected by the performance of the intelligent machines and quality of M2M communication. Thus, in Paper A [57], we argue for merging the conventional QoE with the QoME, and forming Quality of IoT Experience (QoIoT), by studying the human-business and business-machine relationships. Essentially QoIoT is a aggregation of metrics that are measured from two different perspectives: (1) QoE, by extracting metrics from the legacy user-centric management models [59, 60, 73]; and, (2) QoME, by understanding machine's behavior and translating the overall machine performance to business-oriented metrics. A high-level overview of QoIoT is illustrated on Figure 1.5, and can be represented as:

Definition 2. (QoIoT) *A metric that aggregates the delivered quality of an IoT service from the perspective of humans and machines. The end-users' experience is evaluated with subjective and objective human factors, while Quality of Machine Experience (QoME) translates the overall machine performance to business-oriented metrics through objective performance indicators.*

Components in an IoT service	*Key Performance Indicators*	*Classification*
Physical	*Hardware type (e.g. sensors, actuators, cameras, lidar), CPU, OS, raw data, labels, memory, display resolution, data-type format, sampling frequency, sensing characteristics, energy consumption, cryptography*	*QoS*
Network	*Jitter, packet-loss, bandwidth, delay, buffer size, Received-Signal-Strength Indicator (RSSI), Reference Signal Received Quality (RSRQ), Signal-to-Noise Ratio (SNR), etc.*	
Application and service specific	*Application type (e.g. video, voice, IoT, web), Application metrics (e.g. bit-error rate, encoding, bitrate, content genre); Subjective human scores (e.g. MOS, mood, enjoyment, emotions, easy-of-use, end-user's characteristics and background), Objective human scores (e.g. HVS)*	*QoE*
Logical system components	*Benchmarking individual data-sources (e.g. anomaly and novelty detection, context-aware pattern recognition), benchmarking predictive models (e.g. condition monitoring, diagnostics, predictive maintenance), benchmarking the impact of autonomous processes on overall service performance (e.g. productivity, safety, reliability)*	*QoME* *QoIoT*

Figure 1.6: Synergies between QoS, QoE, QoME, and QoIoT.

Figure 1.5 present the relationships between QoE and QoME in forming QoIoT. In an industrial settings, it is the business entity (e.g. stakeholders, service providers and owners) that make decisions to install, replace, and update the machine entity (e.g. IoT objects, M2M communication, AI systems). In addition, the business entity, driven by customers and end-users needs, defines the objective QoME metrics that benchmark the performance of the machine entity. QoIoT aggregates the conventional QoE metrics with the QoME metrics, with a scope to merge the human and machine perspectives in improving the human-business and business-machine relationships. Endsley [24] defines the term Situation Awareness (SA) in dynamic systems; it means to observe and comprehend the current state of the system in the environment, detect patterns and create knowledge that improves service control and performance. Therefore, in complex IoT services services it is essential to create QoIoT knowledge understandable for both entities, human and machines (Figure 1.5). On the one hand, the Human entity needs such knowledge to: (1) Understand and measure the impact of the machines on the business and end-users' QoE, enabling the stakeholder to dynamically adjust and prioritize certain QoME metrics; (2) Replace or install new equipment, and upgrade software; (3) Semi-control the machines; and (4) Improving the end-users' SA of the system, allowing the end-user to promptly react in overriding an autonomous machine or adjusting its settings. On the other hand, measuring QoIoT is beneficial for the Machine entity in understanding the impact of individual data-sources and AI logical components on the overall service performance [77], and finally, making use of the QoIoT metrics as benchmarks for self-learning methods within the autonomous operations.

Figure 1.6 gives an example of conceptual Key Performance Indicators (KPIs) that can lead to QoIoT evaluation, applicable to a general IoT case-study. It also summarizes

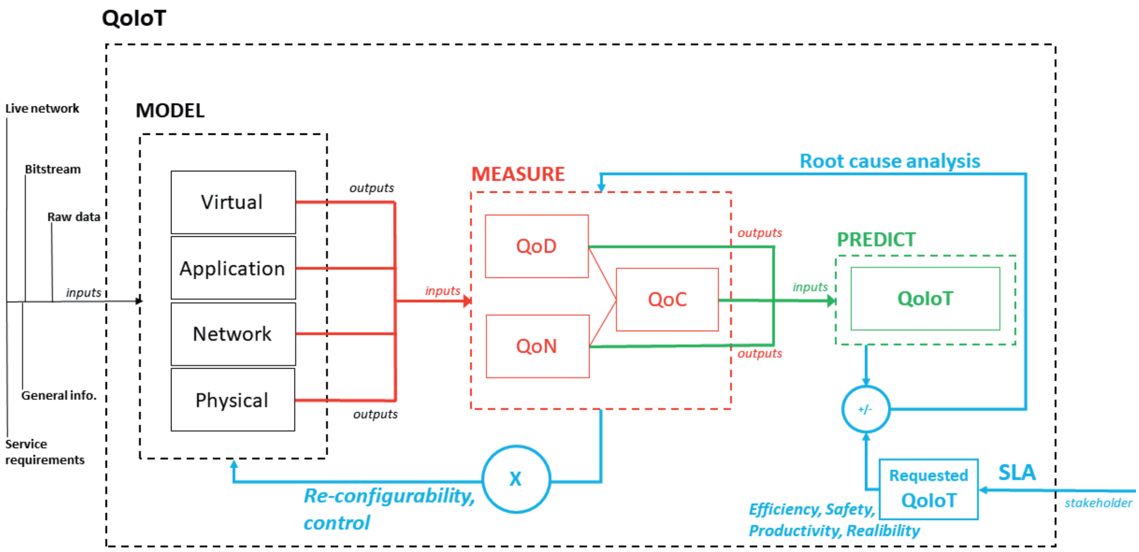


Figure 1.7: QoIoT model, measure, and predict.

the synergies among the current state of the art and QoIoT, by classifying the KPIs based on the main components of an IoT service (e.g. physical, network, application/service, and logical system components). The way QoIoT metrics are aggregated from QoE and QoME will depend on the business goals and service requirements; thus, the business entity (e.g. stakeholders) is the prime actor in their formation. In addition, the QoIoT metrics will depend on the level of machine intelligence, network usage, presence of end-users, and their interactions within the service.

1.5.2 *Contribution 2:* Developing an architecture to model Quality of IoT experience

As a second contribution of this thesis we extend the work presented in Paper A [57] by developing an architecture to model, measure, and predict QoIoT, addressing the RQ2. Having these three steps of the architecture is eliciting a layered-separation of the roles within the quality evaluation process, beneficial when conducting a root cause analysis. With Paper B [55], we explore more in depth the complexity of autonomous vehicles in industrial settings, which we take as a case-study to apply the QoIoT architecture. Figure 1.7 illustrates the proposed architecture and in the following we give an overall description of its main components.

Figure 1.7 shows the prerequisites of input parameters for the model, including both static and dynamic information. They include live access to the raw data, bitstream, and network as a dynamic information, such as data from generated from the IoT objects, logical system components, as well as their communication; while static information is considered to be the general service information, such as specification of the utilized hardware and software, as well as the service requirements. The static information is assumed to be already available in the standardized format, ready for further analysis. The *QoIoT model* consists of four layers (physical, network, application, and virtual), each with a predefined scope and set of methods, which can be defined as follows: (1) Retrieve and understand the static service information; (2) Detect the components and main data-sources of the service within each layer; (3) Monitor the state of the components by gathering the dynamic, real-time data, and prepare measurable metrics that can be used for further analysis; (4) Control and re-configure the physical, network, and virtual components.

The data acquired from the layered model enables transforming low-level metrics into context-aware analytics by adding ML and knowledge-driven approaches supported by software-based platforms. *Measure QoIoT* essentially extracts the data analysis process from the object perception layers in the model. As proposed with the QoIoT definition in Section IV, this entity aims to evaluate the quality of an AV from the human and machine perspective. The main goal is to propose KPIs that are understandable for both humans and machines, improving their SA and enhance decision-making processes. To do that, *Measure QoIoT* proposes three sub-layers that consist of context-aware pattern recognition, novelty, and anomaly detection techniques to evaluate the data generated by the model. Quality of Data (QoD) mainly addresses the metrics generated by the Physical layer, with the aim to test the accuracy of the hardware, but also to find anomalies and discover novel states. Quality of Network (QoN) facilitates situation awareness projection of the way networks impact the overall service performance through enhanced network monitoring features, orchestration of virtual functions, prediction algorithms, and pattern recognition strategies. Quality of Context (QoC) processes the metrics from the previous layers and groups them in greater context - QoE and QoME. The scope of QoC is to understand how QoE and QoME can leverage each other in order to improve the overall service performance. Finally, QoC discusses the formation of the objective knowledge that evaluates QoIoT.

A prediction on the overall QoIoT, as a combination of QoE and QoME KPIs, such as productivity, efficiency, safety, MOS, and technology acceptance, concludes the proposed concept. Figure 1.7 depicts the role of the stakeholder in defining the SLAs by requesting thresholds for each of the measured KPIs. The stakeholder also defines the importance of each KPIs by selecting weights of how much they contribute to the overall QoIoT. This is a dynamic process that can change in real-time, as the stakeholder can alter the quality levels of a certain KPI(s), and thus, QoIoT must adjust to such a scenario. The purpose of the *Predict QoIoT* is to forecast in the near future values of the objective KPIs by utilizing recorded time-series metrics. The idea is to prepare the system and the self-learning methods for the upcoming states and conditions. QoIoT, in its core, supports a feedback loop in a form of root cause analysis on the estimated objective metrics. This is enabled by the structured decoupled layered model, from where a point or contextual anomaly can be traced and isolated back to its origins. In QoIoT, a top-down approach would monitor the QoE and QoME KPIs, such as productivity, efficiency, and safety, and in case of poor indicators, a root cause analysis crawls back through the antecedent layers to identify the outliers in the available data.

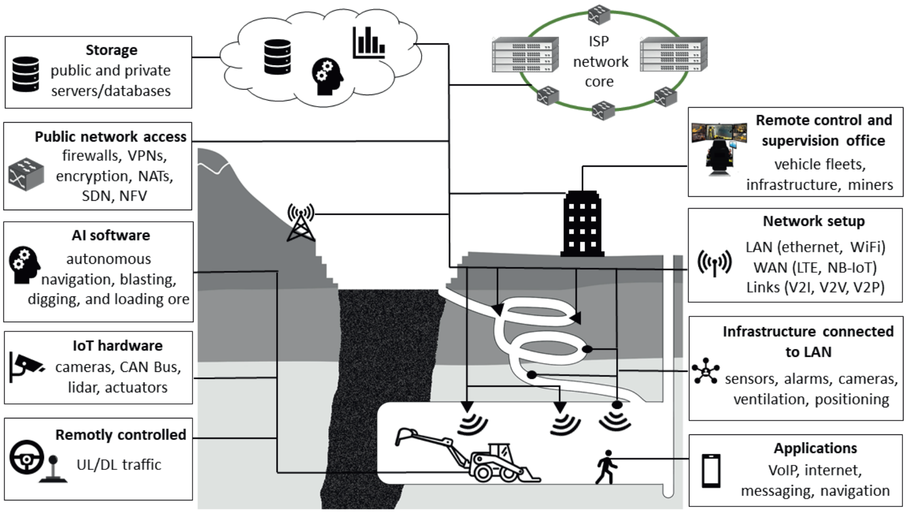


Figure 1.8: The mining experimental setup.

1.5.3 *Contribution 3:* Validating the proposed architecture for Quality of IoT experience

The third, and final, contribution of this thesis, addressing the RQ2, consists of processes for validating the proposed QoIoT architecture from Figure 1.2. For this purpose, real-life experiments were conducted in industrial settings, involving underground mining vehicles. The experimental setup is illustrated on Figure 1.8, where a mining vehicle is remotely-controlled from distance via the internet. The mining vehicle continuously streams video and sensory data (e.g. engine control unit, speed, steering angle, weight, fuel, transmission, antilock braking) to the control room. The data stream travels from the mine, through the public LTE network, with final destination to a remote control station, from where a expert driver is sending steering commands back to the vehicles. The validation of the QoIoT architecture, in this thesis, is first aimed towards the Quality of Network (QoN) layer, namely by automating the QoN process using machine learning, conducting a root-cause analysis, and evaluating the QoN impact on the end-user's QoE.

Real-time estimations of network latency

In Paper C [84] we target real-time estimations of network delay using machine learning approaches. Within the mining IoT setup on Figure 1.8, vehicles drive in and out of the mine, being connected both on conventional cellular network and isolated mining cellular network. Network delay in such case can occur as a results of handovers, interferences, network congestions, and cell edges. The scope of Paper C is to automate the process of measuring the end-to-end network delay, estimating the delay on a millisecond level basis by utilizing passive measurements. We model variety of wireless access link characteristics as features, such as Received Signal Strength Indicator (RSSI), Signal to Interference+Noise Ratio(SINR), and Reference Signal Received Quality (RSRQ), which we measure at the source-node streaming the IoT data. The delay is calculated by taking the timestamp of each packet arriving at the end-node, subtracted with the timestamp of

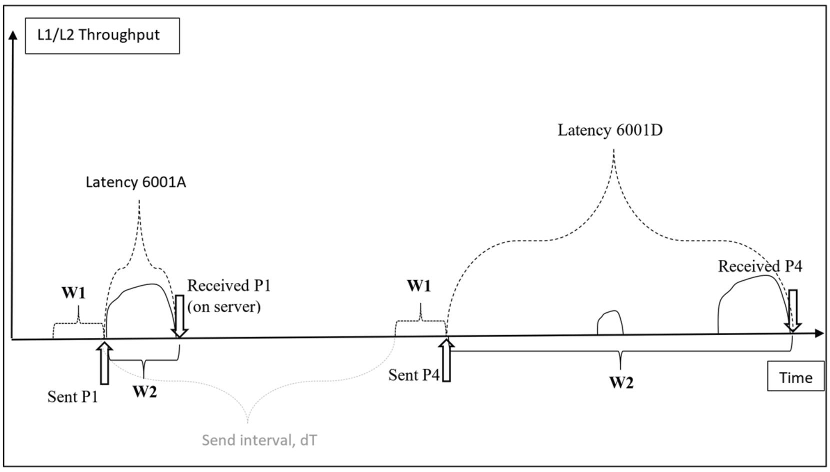


Figure 1.9: Feature engineering for estimating the delay.

the packet leaving the source-node. Then, we perform the experiments to train the features against measured delay values by using neural networks and two types of decision trees. The results of estimating the delay are shown on the Table 1.1.

In practice, the features were collected at the same source-node streaming the IoT data, during a pre-defined sampling frequency period. The process of feature engineering was described within an European patent application (No. EP18306881.6), as a follow up to the Paper C [84]. Figure 1.9 illustrate the process, which consists of two packets (P1 and P4) with the measured corresponding delays (6001A and 6001D). Window W1 is proposed to capture the radio conditions before each scheduled transmission starts, with a varying size depending on the use-case. While window W2 describes the radio conditions after each transmission starts, with a dynamic size depending on the radio conditions and the total transmission time. For instance, the size of W2 for the latency 6001D shall be increased until the following criteria becomes true:

$$sum\_of\_throughput > sum\_of\_packet\_sequence\_size \tag{1.1}$$

where the *sum_of_throughput* is the utilized physical layer throughput since the start

Table 1.1: Latency prediction performance.

	R^2 (coefficient of determination)			Mean Absolute Error (MAE)		
Sampling period	NN (MLP)	Decision Tree	Bagging Decision Tree	NN (MLP)	Decision Tree	Bagging Decision Tree
20 ms	82 %	82.2 %	90.7 %	0.23	0.11	0.091
50 ms	75.5 %	77.7 %	85.1 %	0.28	0.16	0.15
100 ms	73.7 %	67.3 %	81.8 %	0.29	0.19	0.13
200 ms	60 %	50 %	66.8 %	0.35	0.31	0.22

of the scheduled transmission and the *sum_of_packet_sequence_size* is the size of the packet sequence that is scheduled for transmission. In the intrusive case, the *sum_of_throughput* may be retrieved from the utilized Physical layer L1 throughput measures. Whereas in the non-intrusive case, the *sum_of_throughput* may be estimated from other non-intrusive Layer L1 and L2 metrics. This estimation may be a machine learning problem in itself [82] or may be calculated by using a formula, as proposed by Akdeniz et al. [5].

Real-time estimations of video QoE

In Paper D [56] we evaluate the impact of the network on end-user's QoE by automating the video MOS measurements using machine learning. In particular, we utilize the same experimental setup as on Figure 1.8 and train the wireless access link characteristics (e.g. RSSI, SINR, RSRQ, RSRP) against pre-calculated metrics computed by using the ITU-T J.247 recommendation [40]. The choice of using J.247 algorithm is three-fold: (1) J.247 computes variety of Video Quality Assessment (VQA) metrics besides MOS, such as Peak signal-to-noise ratio (PSNR), Frame delay, Frame skips, and Blurriness. Although these kinds of objective metrics do not take into account the characteristics of the HVS in detail, they are relatively simple to compute, giving an acceptable performance because image characteristic is modeled in a perceptual manner [81]. We assert that developing a separate ML model to predict each of those metric using the same data-sets will give versatility to the VQA, broadening the scope of the study; (2) The service requirement by the mining company to utilize J.247, due to licensing and their previous experiences, as well as inability to conduct a subjective surveys with the miners; (3) One goal of the study is to analyze whether the measured wireless access link characteristics can successfully estimate and follow the rate of change in variety of VQA metrics during the mining drive tests.

We assert that the overall goal is to conduct a non-intrusive estimations, using publicly available features, of VQA metrics during a real-time video streaming. In practice, ten video files, with parameters shown on Figure 1.2, were interchangeably used as samples during the streaming process, resulting with a pool of 200 videos for training and testing purposes. The achieved results show 94%, 89%, 84%, 85%, and 82% classification accuracy using decision trees when predicting MOS, PSNR, Frame delay, Frame Skips,

Parameter	Value
GOP length	30 frames
Frame rate	30 fps
Bitrate (avg)	3 Mbps
Resolution	HD 720x480p
Color mode	YUV (4:2:0)
Codec	H.264
Protocol	RTP

Table 1.2: Video encoding and transmission parameters.

	MOS	PSNR	Frame Delay	Frame Skips	Blurriness
R^2	*0.94*	*0.89*	*0.84*	*0.85*	*0.82*
RMSE	*0.192*	*0.236*	*0.317*	*0.298*	*0.359*
		Feature Importance			
RSRP	*0.197*	*0.187*	*0.278*	*0.210*	*0.144*
Cell Power	*0.043*	*0.063*	*0.054*	*0.073*	*0,087*
RSRQ	*0.293*	*0.193*	*0.369*	*0.283*	*0.271*
CQI	*0.123*	*0.083*	*0.064*	*0.013*	*0.034*
MSC	*0.012*	*0.051*	*0.012*	*0.108*	*0.066*
CINR	*0.178*	*0.128*	*0.109*	*0.149*	*0.173*
RS Power	*0.021*	*0.092*	*0.030*	*0.016*	*0.127*
RSSI	*0.133*	*0.176*	*0.174*	*0.148*	*0.098*

(a) Random Forrest classifier results.

Metrics	MOS > 3.5
RSRP	*-98.3 < RSRP > -93.9*
SSS Cell Power	*17.1 < SSS Cell Power > 18.4*
RSRQ	*-9.3 < RSRQ > -8.4*
CQI	*12.2 < CQI > 13.5*
MSC	*18.4 < MSC > 20.2*
CINR	*13.3 < CINR > 14.5*
RS Power	*-76.3 < RS Power > -64.7*
RSSI	*-70.2 < RSSI > -63.9*

(b) Minimal range of values for MOS > 3.5.

Figure 1.10: Results from Paper D [56].

and Blurriness respectively, as shown on Figure 1.10a. In addition, Figure 1.10b shows the range of values for each feature in order to achieve a MOS value above 3.5. Figure 1.10b may be seen as a QoN root-cause analysis, from where few conclusions emerge. For instance, measured RSSI value below -70 dB would advocate bad coverage areas or blind spots, while SINR value below 13 suggests interference and therefore produces poor video quality. Moreover, similar RSRQ values of the serving and neighboring cell indicates cell edge and therefore a handover may be expected, which might cut the video streaming for a short time.

1.6 Chapter Summary

This chapter gives an introduction to the main topic of the thesis, followed by detailed discussion of the research motivation. Therein, we discuss the main research challenges that we identify from the state of the art in the area of QoE and IoT, which we use to motivate and set the scope for this thesis. Further, we define and describe two research questions that we attempt to address with this thesis, centered around defining quality of experience for IoT services and proposing an architecture for its evaluation. Then, we describe the research methodology that we utilized for this thesis, which consist three main milestones: defining research questions, theoretically model the solution, and implement the technical solution. Finally, we list the main thesis contributions that emerged from our research, with detailed discussion on the way each thesis contribution addresses the proposed research questions.

Chapter 2

Background and Related Work

In this chapter we describe the background of the technological concept used in this thesis, as well as discussing in greater details the relevant related works. The Section 2.1 presents the fundamentals of the IoT evolution, coupled with the discussion of the main IoT building blocks. A special attention is given to the area of IoT connectivity, giving an overview of cellular network infrastructure with measurable metrics that we utilize in our third thesis contribution, discussed in Section 1.5.3. Then, in Section 2.2 we cover the fundamentals of QoE, with the most prominent definitions and architectures. Therein, we also discuss the state of the art on QoE for IoT, by analyzing the related works.

2.1 Internet of Things (IoT)

The concept of IoT originates from early 00s with studies on RFID technology by the MIT Auto-ID Center research group [11], which later led to the first architecture designs and applications for the IoT [52]. Xu et al. [20] did a comprehensive study on the IoT evolution throughout the years, discussing the differences from RFIDs and Wireless Sensors Networks (WSNs), to Cyber Physical Systems, and Industrial IoT. What unifies these concepts are the fundamental building blocks shown on Figure 2.1, which structure stayed the same throughout the evolution, where each new developed technology within a block extends its functionalities [12]. For instance, within the communication block, RFID directly paired two devices, and its successor WSNs interconnected multiple devices with a specific protocol, such as ZigBee [20]; while IoT additionally extended the communication block by adding internet connectivity for each device [67]. The current trends in IoT research keep the same track, extending the functionality of the fundamental building blocks, namely by: (1) Exploring opportunities and use-cases for novel access network technologies, such as long and short range communications, using licensed and un-licensed bands [54,67]; (2) Developing new modular boards capable of connecting range of different devices [6]; (3) Proposing novel management methods for governing an IoT service [23]; and (4) Making the whole IoT eco-system more secure [52].

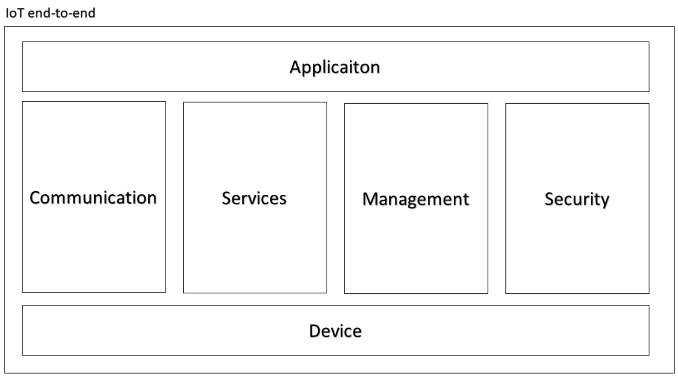


Figure 2.1: The IoT building block.

2.1.1 IoT End-to-end Services

The research advancements in the IoT eco-system facilitate novel applications and services in solving some of the most challenging quests. The most recent use-cases include energy management systems [53], remote-controlled surgeries [69], sustainable rural development [21], and mitigating natural disasters [32]. In the following, we present an overview of the state-of-the art in IoT, with respect to the the IoT building blocks (Figure 2.1), that enable the emerging end-to-end IoT services.

Device. Represent physical objects which can perform perception tasks, scanning and sensing their surrounding environment [46]. Few examples include sensing temperature, motion, and vibration [46]. In addition, IoT physical objects can be in a form of actuators, waiting for a signal to trigger and execute processes [6]. Generally, the IoT sensors and actuators aim to collect, process, and share information [6]. In recent years a modular set of hardware emerged, under the name of Single Board Computers (SBCs), such as Arduino and Raspberry Pi, offering computational power, storage and internet access. Such hardware is modular, with capabilities of plugging multiple sensors and actuators simultaneously, via a range of ports (e.g. digital, analog, USB, and HDMI), that can collectively be managed by an operating system [6]. IoT devices are envisioned to be plug-and-play devices [6], where the possibility to access the internet opens up various vulnerabilities, summarized by Sharma et al. [70]. Many of the challenges, such as security, management, or connectivity issues, can be address by configuring the devices with the latest standardized approaches, surveyed by Al et al. [70]. However, a fundamental issue within the device block, as pointed out by [18], arise from various hardware and software failures, affecting the devices' sensing as well as computation, which we discussed in the Section 1.2 and attempt to address with the second thesis contribution, in Section 1.5.2.

Communication. The communication block is the key enabler for IoT services as it gives the physical objects possibility to access the internet. The continuous evolution of the communication block enables the pervasiveness of the IoT as an technology, for both private and business use [12]. Al et al. [6] presents an overview of the IoT communication block, where IoT-tailored network and application protocols are reviewed from the perspective of service requirements and IoT devices' capabilities. For instance, a SBC can be configured with antennas capable of short or long-range network protocols, such as Bluetooth or NB-IoT, on top of which can be arranged one or multiple application protocols, such as CoAP, MQTT, and XMPP [6], depending on the service requirements. A study by Zhang [83] reviews the IoT communication protocols from a perspective of the frequency bands. A high-level classification divides the network protocols operating in licensed and un-licenses frequency bands. Zheng et al. [83] discuss the advantages and disadvantages of both classes, pointing out that unlicensed protocols, such as LoRa, SigFox, and Bluetooth are inexpensive, but offer limited or no quality guarantees. On the other side, the licensed protocols are more expensive, however, allow access to the legacy cellular network infrastructure that offers dedicated channels with quality guarantees [83]. The scope of this thesis is centered around industrial IoT case-studies, described in the Section 1.5.2, which pose high demands on the network quality and thus, the work presented in this thesis considers cellular network communication protocols. Due to the high importance of the communication block, in the following Section 2.1.2, we will provide more details on the IoT cellular connectivity, reviewing the state of the art and identifying the key challenges that we attempt to address with the contributions in the Sections 1.5.2 and 1.5.3.

Management. Management in an end-to-end IoT service can refer to device, data, and network management [6], which are respectively described in the following: (1) Multiple software platforms exist that offer IoT *device management*, typically deployed on a centralized SBC. There are several real-time operating systems that are good candidates for managing IoT devices, such as Contiki, Tiny OS, and Android, which are reviewed by Al et al. [6]; (2) The produced raw data by the IoT devices is typically handled with *data management* techniques that can be deployed on the local hardware or accessed online. Data management platforms offer the possibilities for graphical representation, monitoring, and advanced data analytics on the generated IoT data [29]. Such platforms are becoming increasingly more useful as they can also perform Machine Leaning (ML) techniques that learn from the available data to predict certain events and states, adding intelligence that can lead into making automated decisions [29]. Gharaibeh et al. [29] reviewed popular commercial data management platforms, by Amazon, Google, and Microsoft, based on their capabilities for data transformation, parameter tuning, creation of new algorithms, and support for various use-cases; (3) *Network management* addresses the challenges that come with massive deployment of IoT devices, such as network scalability and capacity [3]. Recent technologies such as Software Defined Networks (SDN) and Network Function Virtualization (NFV) enable network management mechanisms with dynamic orchestration and programmable resource allocation, which can significantly improve the overall service performance [3]. When developing the proposed QoIoT architecture, described in Section 1.5.2, we consider quantifying the functionalities of the management block in evaluating their impact on the overall IoT service quality. However, the validation of the management block within QoIoT is left for the future work.

Services. A fundamental characteristic of an IoT service, especially in the industrial domain, is the coalescence with functionalities provided by external services [46]; this has led to creating complex service chains, known as System-of-systems (SoS) [22], with possibly even hundreds of services offered by different third parties, each with their own business incentives. Consider an IoT service where a prerequisite of executing a certain process is an input data coming from an external service, such as a weather forecast [46]. Such service dependencies are not unique just for IoT scenarios and typically are handled by a signed SLA agreements, where certain quality requirements must be delivered [46]. In a typical case, a service share the IoT generated data through IoT cloud platforms, which offer the possibility to publish the data to any subscribed external services in a standardized format [22].

In this thesis, within the proposed QoIoT architecture, we raise the importance of considering the described IoT building blocks when evaluation the overall IoT service quality. However, the validation of the QoIoT architecture focuses on parts of the device and communication blocks, with the Papers C and D [55, 57]. We assert that validating the management and service block within QoIoT is currently out of scope and, thus, left for future work.

2.1.2 IoT Cellular Connectivity

The proliferation of IoT services in the industry depends, among other, on reliable performance of the IoT connectivity [6]. For instance, IoT use-cases for traffic safety, industrial manufacturing, and remote-controlling (e.g. health-care, transportation) are considered as mission-critical services, where the quality of the network communication may lead to endangering human lives, environment, and business [46]. Thus, as mentioned before, the IoT service providers turn into the legacy cellular networks to get reliable connectivity. Compared to the unlicnesed IoT protocols (e.g. LoRa, Sigfox), the legacy cellular networks can offer guaranteed network performance regarding achieved throughput, minimal latencies, jitter, and packet losses [83], with characteristics illustrated on Figure 2.2. In the following, we will explore the state of the art in the cellular network infrastructure, as well as provide the key challenges in cellular IoT.

Background Work on Cellular Network IoT Infrastructure

Current cellular network technologies, such as GSM, 3G, and LTE, are not designed to simultaneously serve massive amount of connected IoT devices [6]. In addition, the emerging IoT services have unique requirements regarding the network connectivity which are vastly different compared to the conventional mobile services, such as VoIP, web-browsing, and streaming [83]. For instance, IoT devices are expected to be intermittently connected to the network, spending most of their time in idle/sleep mode, using a fixed and low transmission rates [66]. Thus, 3GPP, as a standardization organization, in its recent releases makes effort to integrate the IoT devices and their requirements in the cellular infrastructure [51]. For instance, 3GPP Release 13, published in 2016, presents design requirements for three IoT protocols: EC-GSM, LTE-M, and NB-IoT [51]. Each of them are suitable for a specific set of IoT use-cases, providing different KPIs, which are illustrated on the Figure 2.2. Generally, EC-GSM is envisioned to operate and replace the legacy GSM bands, while LTE-M and NB-IoT will share the resources with the legacy LTE technology and its frequency bands [51]. In addition, NB-IoT will have three deployment modes, where in can co-exist with LTE and LTE-M within the same bands, but also be deployed between the bands and in guard bands, with respect to time [51]. Prior knowledge of the characteristics of the network protocols is necessary when modeling the network quality in an IoT service, and hence, within the Section 1.5.2, we define a network layer that considers the service requirements for the network protocols.

Parameter	Non-cellular network		Optimization of IoT capabilities		Cellular network
	LoRa	*Sigfox*	*EC-GSM*	*LTE-M*	*NB-IoT*
Range	<15 *km*	<50 *km*	<15 *km*	<11 *km*	<15 *km*
Maximum Coupling Loss	157 *dB*	153 *dB*	164 *dB*	160 *dB*	164 *dB*
Maximun Peak Data Rate	50 *kbps*	100 *bps*	74 *kbps*	1 *Mbps*	250 *kbps*
Average Module Cost	7~10 dollars	Less than 5 dollars	5 dollars	Less than 10 dollars	Less than 5 dollars
Devices	10,000	10,000	20,000	18,000	50,000
Spectrum	Unlicensed EU 868, 433 *MHz* US 915 *MHz*	Unlicensed EU 868~869 *MHz* US 902~928 *MHz*	GSM Bands	LTE Bands	LTE In-band Guard-band Stand-alone
Bandwidth	<500 *kHz*	100 *kHz*	200 *kHz/ch.*	1.8 *MHz*	180 *kHz*
Radio Technology	Spread Spectrum	Ultra Narrow Band	TDMA/FDMA	OFDM	OFDM
Voice	No	No	No	Yes	No
Autonomy	>10 years	>10 years	>10 years	>10 years	>10 years
Lnk Aaption	Yes	No	Yes	Yes	Yes
Industry support and maturity	Commercial		Plan	Supported	Commercial

Figure 2.2: Performance indicatiors for the IoT protocols.

Several recent research studies see the characteristics of the protocols from Figure 2.2 as insufficient in covering mission critical IoT services [46, 51, 83]. Consider remotely-controlled vehicles as a use-case, described in the Section 1.5.3, with abundance of delay sensitive packets streamed in uplink and downlink, including live-video (uplink), vehicle's monitoring systems and sensors (uplink), and real-time control stream (downlink). The requirements, as defined by 3GPP, are end-to-end latency of less than 50ms with block error rate (BLER) of 10^{-5} [2], with gigabit throughput values for the video stream, opposed with low values for the real-time control stream. Thus, 3GPP initiated the works on extending the 4G/LTE network with requirements for the upcoming 5G/NR, envisioned to enable the mission critical IoT services [66]. To do so, 3GPP defined requirements for end-to-end network slicing, which will horizontally slice computing and communication resources to form virtual interfaces for supporting vertical industry applications with diverse requirements, e.g., in terms of functionality, performance, and isolation [2]. Concepts such as SDN and NFV are key for enabling network slicing within the 5G/NR, which we consider in our architecture for modeling the quality of an IoT service, described in the Section 1.5.2.

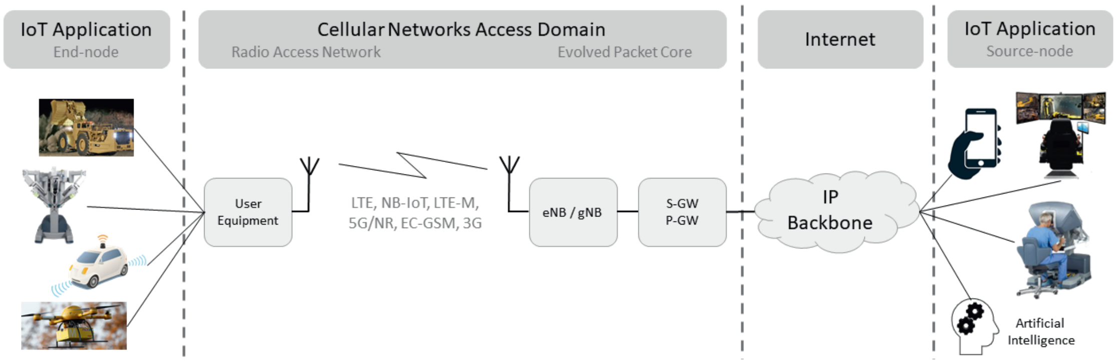


Figure 2.3: Cellular network IoT infrastructure.

Evaluating the Wireless Network Access Link

Figure 2.3 illustrates an overview of the cellular network IoT infrastructure, envisioned to be backwards compatible and capable of serving devices from any of the legacy licensed network protocols [6]. Sharma et al. [70] discuss the significance of compatibility among technologies, especially in dynamic, vehicular cases, where a mobile device can switch from LTE to 3G and GSM due to network coverage constrains. As mentioned before, the performance levels of the network is of a upmost importance for businesses investing in IoT services [46], especially when it comes to mission critical IoT that can impact human lives. Thus, the contributions proposed in the Section 1.5.3 of this thesis, containing Paper C [84] and Paper D [56], considers evaluating the impact of the cellular link protocols on the certain network quality parameters.

Real-time evaluation of the wireless network access link is typically done by inspecting the packet-header and encoded bitstream, e.g., sequence numbers, RTP markers, as suggested by ITU-T Y.1540 recommendation [45]. In addition, different kind of probing messages are recommended by [45], such as sending a ping or unbiased message sample to calculate delay and jitter, as one of the network quality parameters. A study by Barakovic et al. [13] reviews multiple frameworks that follow the ITU-T guidelines for creating end-to-end network evaluation that utilize packet inspection and bistream information. However, Robitza et al. [68] pointed out that the service providers are recently encrypting their traffic, which limits the access to those evaluation metrics. More complex network evaluation systems utilize bulk transfer tools to calculate network quality metrics, such as throughput, delay, jitter, and packet losses [47, 78]. For instance, SCReAM [47] and Coupled CC [78] measure congestion in the network by streaming the maximum amount of allowed traffic during an observation windows. However, having an observation window is a limiting option for most of the IoT use-cases, due to multiple constrains, such as power/battery usage, scarce resources, and dynamic environment, which requires constant measurements. More cost-effective techniques involve monitoring passive, non-intrusive, metrics that can be obtained by observing the wireless link between the devices and the base-stations, as illustrated on Figure 2.3. Afroz et al. [4] discuss the strong correlation between non-intrusive metrics, such as RSSI, RSRQ, RSRP, with events such as available throughput, link adaptation, handovers, and packet scheduling.

Feature	Description	Mode
RSSI	*Carrier Received Signal Strength Indicator: comprises the linear average of the total received power (in W) observed only in OFDM reference symbols [1]*	Idle
RSRP	*Reference Signal Received Power: the linear average over the power contributions (in W) of the resource elements that carry cell-specific reference signals [1]*	Idle
RSRQ	*Reference Signal Received Quality: the ratio N*RSRP/(Carrier RSSI), where N is the number of resource blocks of the carrier RSSI measurement bandwidth [1]*	Idle
SSS Cell Power	*Secondary Synchronization Signal power for detected cells (in dB) [1]*	Idle
Total RS Power	*Total Reference Signal power calculated from serving cell RSRP and channel bandwidth [1]*	Idle
CQI	*Channel Quality Indicator for code word 0 [1]*	Conn.
PUSCH MSC	*Modulation Coding Scheme index for the uplink transport block [1]*	Conn.
RS CINR	*Carrier to Interference plus Noise Ratio of the signal carrier best servings for the intervention seemed at all other sites/sectors, plus all the noise*	Conn.

Table 2.1: Wireless access link characteristics.

Table 2.1 presents the description of multiple wireless access link characteristics that we utilize within our Papers C and D [56, 84], which are retrieved by scanning the radio and analyzing the signals from the surrounding cells (Figure 2.3), part of the 3G, LTE, 5G/NR, and their IoT versions [1, 51].

Challenges in Cellular IoT

The 3GPP's efforts to design new IoT network protocols can mitigate some of the challenges regarding the deployment of massive amount of IoT devices, extended battery life, production of low-cost devices, as well as achieve extended radio coverage [51]. However, the introduction of IoT communication within the cellular network is challenging due to the appearance of high interference with the legacy protocols [62]. Oh et al. [62] points out at the interference created within the LTE bands once IoT protocols such as NB-IoT and LTE-M are commercially deployed, resulting with failures in the communication. In similar way, high interference is anticipated in the new 5G/NR structure, where an IoT slice is characterized with low-throughout and ultra low latency [66]. Popovski et al. [66] illustrates the cellular resources given to an IoT slice, which take short time-slots, but large portion of the free resources per time-slot and such schema will create high interfere with the conventional human-type communication (e.g. voice, messaging, web-browsing, video streaming, and gaming). Thus, with the second and third contribution of this thesis, discussed in Sections 1.5.2 and 1.5.3, we consider evaluating the impact of the network on the end-user perspective within an IoT service.

Another challenge created by the proliferation of IoT is the machine-to-machine (M2M) communication, where multiple IoT devices can communicate among themselves to produce an outcome. Section 1.2 describes the case of self-driving, where autonomous decision are made based solely on the M2M communication. As pointed out by Van et al. [77], there is a challenge in evaluating the impact of the M2M communication on the overall autonomous system, due to the absence of control algorithms that are necessary to benchmark the performance of the autonomous decision [77]. The presence of end-users within the autonomous services additionally complicates the quality evaluation process, which we consider with the first contribution of the thesis, described in Section 1.5.1. However, before discussing any details on the end-user's perception of the M2M and autonomous systems, due to the complexity of the topic [77], a thorough discussion is first necessary on the state of the art within the subject of QoE, which follows in the next section.

2.2 Quality of Experience (QoE)

This section is divided into two parts, first, to capture the QoE fundamentals and to model, measure, and predict QoE; and secondly, to show how the state of the art QoE concept differs when it comes to evaluating QoE for IoT services.

2.2.1 QoE Fundamentals

The understanding of the term *quality* and its importance for the Human-Computer Interaction (HCI) led both the academic and industrial sectors into developing concepts such as Quality of Service (QoS), User Experience (UE), and Quality of Experience (QoE). QoS is well-established area with abundant research [13, 59, 73], where the focus is on *what* is happening in the network by investigating objective metrics such as packet loss, latency, jitter, and throughput [59]. Thus, QoS is typically utilized by network providers when building, upgrading, and maintaining their infrastructure - composed of core networks, base stations, routers, switches, and the logical links among them [13]. However, typical QoS model fail to provide accurate insight in the *why*-dimension [59]. For instance, describing "why is the user behaving in a certain way?" and "why does the user feel frustrated?". The user does not perceive the individual network element but rather feels the overall system performance [27]. To highlight the user point of view, a new concept was introduced, the QoE. According to the ITU-T [39], QoE is defined as: *"The overall acceptability of an application or service, as perceived subjectively by the end-user."* It is worth noting that ITU-T also considers the following statements: *"Quality of Experience includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc.)"* and *"Overall acceptability may be influenced by user expectations and context."* A key noticeable point in those definitions is that ITU-T does

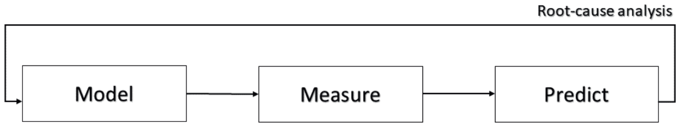


Figure 2.4: Model, Measure, and Predict QoE.

not describe what it means by *context* and how experts can measure users' *expectations*. Therefore, our conception of QoE is in line with the work of [59], who states that *"Quality of experience(QoE) is a metric that depends on the underlying QoS along with a person's preferences towards a particular object or service where his/her preferences are defined by his/her personal attributes related to expectations, experiences, behavior, cognitive abilities, object's attributes and the environment surrounding that person"*. Moreover, this thesis is based on grouping the concept behind the evaluation of QoE by following the three steps - model, measure and predict, as suggested in [59] and depicted on the Figure 2.4.

2.2.1.1 Model QoE

The first set of QoE models found in the literature are in a conceptual form [60]. For instance, a comprehensive survey by Alreshoody et al. [19] identifies parameters that can impact the overall QoE, where their classification is divided into four groups: (1) The standard QoS parameters, such as packet loss, delay, jitter, and throughput; (2) Subjective parameters related to the application and network, such as enjoyment, mood, and emotions; (3) Usability parameters such as easy-of-use and technology acceptance; and (4) Context parameters such as personal, cultural, social, environmental, and technological. The conceptual models then transitioned into detailed methodological steps for measuring and predicting users' QoE in both laboratory and real-life environment. More recent models focus on the relationships among the discussed parameters, describing the impact of, for example, QoS parameters on usability parameters. For instance, Alreshoodi et al. [7] surveys the correlation analysis methods among QoE metrics, then Mitra et al. [59] uses Bayesian Networks (BNs) to measure subjective QoE ratings regarding different context situations. In this case, BNs handle the both discrete and continuous parameters, as well as their linear and non-linear relationships. However, the literature on QoE is rather diverse and scattered across fields. The inconsistencies are mainly at the level and amount of factors that influence users' experience [59]. Also, the relationships among the measured parameters is quite complex due to their abundance [60]. For instance, there can be *1, ..., n* contextual parameters affecting *1, ..., m* subjective parameters. Further, n contextual parameters can affect each other based on the nature of the application and service. Thus, researchers propose methods of selecting specific QoE metrics based on the end-user target group, type of service, cultural, and social preferences [19]. However, the state of the art QoE models mainly cover multimedia services, such as video, voice, and web [19]. The advent of new multimedia compression and codec standards, new transmission system and consuming technologies, all call for a new and better understanding of QoE [19].

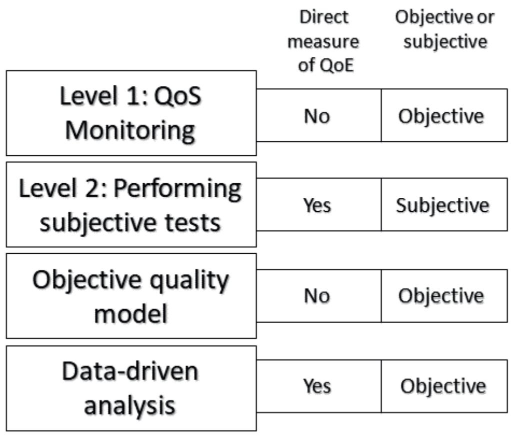


Figure 2.5: Evolution of QoE evaluation.

2.2.1.2 Measure QoE

The general separation when discussing measurement QoE techniques is divided among *objective*, the rationalistic and product-oriented approach, and *subjective*, the user-oriented perceptual approach [59]. A study by Chen et al. [19] maps the linear evolution of the QoE measuring process, illustrated on the Figure 2.5, which can be classified as:

- **Subjective.** Quantitative assessment of, for example, audio and video quality inevitably involves human observers participating in controlled laboratory tests, which elicit quality scale ratings [60]. ITU-T with P.800 recommendation [38] presents a methodology for conducting subjective tests and plotting the results on a rating scale. Though subjective QoE is viewed as relatively accurate, there are number of drawbacks associated with it: (1) They require large sample space for credible results, making them costly in terms of time and resources [60]; (2) Laboratory tests limits the application/service type, tests conditions, and viewing demography, making them inaccurate for real-life QoE assessment; and (3) Subjective tests cannot be used in real-time QoE evaluation [19].

- **Objective.** The objective quality assessment compute a metric as a function of QoS and the way Human Visual System (HVS) perceive and process the information [19]. The most commonly used QoE objective models, such as E-Model [42], PEVQ [40], and POLQA [44], then verify their objectively computed score with subjective tests. Offline models, such as [40,42,44], typically stream a reference file to a end-node, and then in post-processing analyze its differences with the original file. For instance, POLQA looks at lost packets, time–frequency mapping, time alignment, alignment with a calibrated listening level, compressed loudness scaling, and frequency warping. However, a disadvantage of such models is the dependency on the original file, which hinders online QoE Prediction [19]. Recently QoE models adopt more advanced data-analysis with a goal to serve as online QoE methods.

2.2.1.3 Predict QoE

Data-driven analytics emerged as the most prominent method of assessing QoE in multimedia [19]. The advancements in statistical and machine learning (ML) methods leverage the disadvantages of the subjective and objective QoE evaluation. Firstly, as a substitute for a subjective rating scales, the abundance of data recorded from, for example, online multimedia platforms can be used to describe end-users' perception through viewing/listening time, number of watched videos, clicks on the page, and context switching. Few data-driven QoE analysis models utilize such metrics to predict the probability for the end-users to return to the platform, or to termite their subscription (e.g. churn) [75, 79]. Secondly, the outcome from offline QoE evaluation models can be utilized as a label towards which a ML model can be constructed. The idea is to record real-time available metrics and train them to predict a QoE score from a rating scale. A popular approach is to use network statistics data retrieved from the packet-header and bitstream. For instance, a study by [10] relies on packet-header information to build a online model capable of predicting QoE in real-time. However, there is a growing trend among popular multimedia platforms to encrypt the packet-header and bitstream information, thus making the QoE models relying on them no longer viable [68]. Going over the encryption issues, ITU-T with the P.1203 recommendation [43] defines modes depending on the encryption levels. Orsolic et al. [63] described the use of ITU-T P.1203 to build various ML models for QoE score prediction on Youtube videos. The study uses in-network features mainly based on calculating mathematical statistics on lists that store the amount of data downloaded by the client in 1 or 5 second periods. ITU-T P.1203 and thus [63] are tailored for HTTP adaptive streaming, targeting issues such as re-constructing missing frames, or buffering of downlink traffic. Although P.1203 recommendation is viewed as a accurate QoE solution for adaptive streaming applications, such as Netflix and Youtube, it is not applicable for real-time streaming services, such as ViLTE, IPTV, Skype, surveillance cameras, or other OTT. Therein, the uplink stream suffers from different issues, mainly frame/packet dropping.

Recently ML-powered QoE models emerged as universal solutions applicable for most of the multimedia application, regardless of the nature of the application (e.g. adaptive or real-time) and security measures. Casas et al [17] collected in-smartphone measurements to predict a subjective QoE of different applications, including Youtube, Facebook, and google maps. Therein, the main features correlated with the subjective QoE are session volume, signal strength, and throughout. Such models are referred as active QoE prediction models, where the predictions take place only during an active data transmission. In Paper D [56], we take a step further in collective passive measurements to predict end-users' QoE [56] for video streaming application. Whilst, Pedras at al. [64] applied similar setup for voice streaming application. The main finding of these two studies is the correlation among LTE radio link characteristics such as RSSI, RSRP, RSRQ, and SINR on the subjective QoE score. The passive nature of these features enables a source-node, e.g. a smartphone, to measure them while being in idle mode and not transmitting any data. LTE radio link characteristics were initially studied by Afroz et al. [4], where their effect was evaluated on objective QoS metrics, such as throughput, link adaptation, packet scheduling, and handover.

2.2.2 QoE for IoT

As shown with the previous sections, the concepts of QoE and IoT are widely discussed in the research community, however, in isolation. The combination of QoE for IoT is still in its infancy due to the very few papers that target this area and lack of clear definition on how emerging IoT services, besides multimedia, impact the QoE. Table 2.2 groups the research studies that target the area of QoE for IoT and classifies them by the following criteria:

- *Application domain.* Related to the overall scope of the study in terms of which IoT applications are covered.
- *Quality of Data.* Related to evaluating the accuracy and integrity of the produced data by the IoT devices, testing their hardware and software. In addition, it includes various way for conducting conditional monitoring of the devices for anomaly and failure detection and isolation, as well as discovering novel states.
- *Quality of Network.* Related to situation awareness projection of the way networks impact the overall service performance through enhanced network monitoring features, such as prediction algorithms and pattern recognition strategies, as well as an evaluation of M2M communication and virtual processes.
- *Quality of Context.* The criteria is whether a study considers and reasons on the contextual situations in which the QoE is being evaluated.
- *IoT View.* Relates to whether a study consider evaluating the quality from the perspective of humans and machines. An outcome can be in a form of metrics that describe the experiences from humans and machine perspective, respectively.
- *Defining QoE for IoT.* The criteria is whether a study extends or re-defines the conceptional QoE concept to cover IoT cases by proposing novel definition. In addition, the definition should cover emerging IoT services that include autonomous processes.
- *Validated.* The criteria is whether the proposed model, architecture, or framework is validated with experimental work.

The study by Wu et at. [80] present one of the earliest works discussing QoE for IoT. They propose the concept of Cognitive IoT as a new network paradigm where physical and virtual things or objects are interconnected and perform as agents. Moreover, Wu et al. [80] sets the ground of QoE for IoT by identifying the layers of data in an IoT service that need to be considered for evaluation. Namely, the authors distinguish a data layer from a information layer, where the first one aims to evaluate the quality of the sensed data, while the later one deals with the intelligent decision-making based on a group of data sources.

Paper (Year)	**Application Domain**	**Quality of Data**	**Quality of Network**	**Quality of Context**	**IoT View**	**Defining QoE for IoT**	**Validated**
Wu et al. [80] (2014)	General	Yes	No	Yes	Human	No	No
Floris and Atzori [28] (2016)	Multimedia	Partially	Partially	Yes	Human	No	Partially
Karaadi et al. [48] (2017)	Multimedia	No	Yes	No	Human	No	No
Shin [72] (2017)	Wearable	No	Yes	No	Human	No	Yes
Huang et al. [37] (2018)	Multimedia	No	Yes	No	Human	No	Yes
Hoffman and Novak [35] (2018)	General	No	No	No	Human and Machine	Yes	No
Suryanegara et al. [76] (2019)	Smart City	No	No	No	Human	Yes	Yes
Minovski et al. [55], Paper B (2020)	General	Yes	Yes	Yes	Human and Machine	Yes	No

Table 2.2: QoE for IoT comparison among studies.

As one of the first studies published in the area of QoE for IoT, Wu et al. [80] identifies few key research challenges that face QoE assessment methods, summarized as: (1) A clear challenge of obtaining high quality data with mixed characteristics and massive amount of sources; (2) Current QoE state of the art is limited on a single user QoE, failing to cover system-level QoE suitable for large scale IoT services with massive user base; and (3) A clear need for developing effective semantic technologies and knowledge discovery suitable for large scale IoT services in understanding their performance. Regarding the Table 2.2 classification, Wu et al. [80] do not consider evaluating the performance of the network on the overall QoE. As mentioned withing the QoN criteria, the study neglect the impact of the communication protocols and radio environment on the service performance, as well as not considering benchmarking the M2M communication among IoT devices and virtualized processes (e.g. cloud computing). In addition, the study do not propose any extension or re-definition of the QoE concept to cover IoT case and just looks at QoE from the human perspective.

A study by Floris and Atzori [28] takes a step further in developing a concept of layered-QoE framework to evaluate the quality of an multimedia IoT service. The validation of the model is performed with gathered data from IoT vehicle application, containing sensors and camera data. The data-sources are streamed and displayed to end-users, which then subjectively assign a QoE rating score. Within their model, Floris and

Atzori [28] propose the concept of Quality of Data (QoD) benchmarking the accuracy of the produced IoT data. Although the authors do not provide methodological steps on what it actually takes to measure QoD, they use an abstracted version of it in combination with QoS to map to subjective QoE scores. Their results are mainly focused on end-user's subjective evaluation, in a form of a MOS score, of the data generated by the system, which includes video streaming and speed/RPM values represented graphically and textually. The authors of the study, however, do not provide any indications of how their framework could cover other services besides multimedia, especially services that involve autonomous processes. In addition, the authors just mentioned few high-level network evaluation metrics, such as delay, jitter, and packet loss, but do do not provide any comprehensive analysis of the network's impact on the service performance, especially important for a vehicle application due to the dynamic wireless environment. Finally, the authors do not try to extend or re-define the QoE concept, but instead follow the legacy QoE by evaluating the subjective end-users opinions.

Karaadi et al. [48] propose the concept of Quality of Things (QoT) that targets the quality of multimedia IoT services, such as automated camera detection of vehicle's number plates and face detection. They identify the communication in future IoT services to be predominantly machine-to-machine (M2M), and thus, motive their work as conventional QoE can no longer be applied in M2M. The QoT architecture contains physical, network, virtual, service, and application layers, with identifying factors affecting the QoT. However, the proposed factors are still on a conceptual level and the discussion fail short in describing the essence of the challenges regarding IoT devices, which are the classification factors in the Table 2.2. For instance, the study by Karaadi et al. [48] discuss the importance of M2M communication, but there is a lack of discussion on cause and effect of chained M2M communication on the overall quality of the produced IoT data. The study also do not consider any contextual information of the proposed study when evaluating QoT, looks only from the human perspective, and fail short in distinguishing the proposed QoT concept from the state of the art in the QoE domain.

Similarly to the previous two studies, Huang et al. [37] focus on multimedia IoT applications, with main objective to find relationship between the multimedia IoT devices and subjective end-users' MOS. The authors consider IoT devices that produce multimedia content, such as video, audio, and still images, from which they store historical multimodal data-sets. Then, neural network-based data fusion technique is utilized to build the mapping between the average MOS and the network related factors, such as bandwidth, delay, loss, and jitter. Based on the classification criteria in Table 2.2, the study by Huang et al. [37] do not attempt to distinguish among network failures and failures generated by the multimedia devices, noted as Quality of Data. In addition, they do not evaluate the QoE by considering any contextual information regarding the IoT use-case. Finally, the authors [37] follow the path of conventional QoE evaluation from the end-users' subjective perception and, thus, do not attempt to extend or re-define the QoE within the IoT.

Hoffman and Novak [35] is one of the first works that attempt to re-defining the QoE for IoT domain, targeting variety of IoT applications, such as smart cities, wearables, and robots. They propose the concept of *object experience*, which does not require smart IoT objects to have conscious experiences in order to have some type of experience [35]. The object experience emerges from all time interactions that involve the object, which can include interaction between the object and other parts, such as consumers and other objects. The authors [35] build a comprehensive theoretical approach to define the object experience, by analyzing its emergent properties, capacities, and expressive roles from all the object-centric interactions. Hence, the object experience understands the role of M2M communication among IoT objects, which can lead to autonomous decisions where humans are not necessarily involved in the decision making process. However, the authors do not provide any measurable metrics of the object experience, and thus do no attempt to validate their approach. In addition, they do not consider any technical details of evaluating the M2M communication among the IoT devices, as well as neglect the implications of the network on the service quality and performance.

Shin [72] proposes a framework for evaluating the end-user perception of a wearable IoT service by finding the relationships among subjective QoE and QoS metric. The study distinguishes among content, system, and service quality, but ultimately reasons on such concepts from the human perspective, by evaluating metrics such as enjoyment and pleasure, associated only with human experiences. Further, Shin [72] conducted subjective interviews to qualitatively understand the end-user's QoE of using the wearable IoT service. Similarly to the previously cited studies, such as [28,37], the main focus is to find correlation among end-users' subjective MOS and QoS metrics (speed, jitter, delay, loss, bandwidth, and burst). The author, therefore, neglect the importance of evaluating the quality of the produced data with measurable metrics, as well as do not consider any contextual information of the IoT use-case and end-users.

Suryanegara et al. [76] propose a 5-step framework for measuring QoE of IoT services, which includes: (1) Setting up the focus of the IoT services and the QoE parameters being assessed; (2) Defining users of IoT services; (3) Conducting a mean opinion score (MOS) survey of the users; (4) Calculating the DMOS (differential MOS) on the basis of the ACR-HR quantitative scale; and (4) Providing the strategic implications to those who implement the service. The first step is important because it sets-up the stage of measurable QoE metrics for IoT services; the authors propose five core parameters, in a form of efficiency, consistency, convenience, integration, functional stability. However, the authors measure those metrics form the perspective of the end-users. That is, a conducted survey asks the end-users to assign a MOS value to each of the proposed metrics based on their subjective experiences of using the service. The authors [76] do not attempt to gather any quantitative data from the service, such as metrics from the IoT devices or network evaluation metrics. Thus, there is a lack of methodological steps that models the data-sources of an IoT service, which can lead into understanding why does the user's assigned certain MOS value. Such quantitative evaluation, in a form of QoD, QoN, and QoC in Table 2.2, is important in cases when IoT services include intelligent machines and autonomous processes due to the ability to conduct a root-cause analysis. In addition, as pointed out by Hoffman and Novak [35] and the Paper B [55], discussed in Section 1.5.2, intelligent machines have their own experience which has to be evaluated.

2.3 Chapter Summary

This section provides a discussion on the background and related work on the two main topics that are covered in this thesis: Internet of Things (IoT) and Quality of Experience (QoE). It starts with a description of the main IoT building blocks, identified by several research studies [12,20], which include device, communication, service, management, and security block. Therein, we point out at the scope of this thesis, with respect to the building blocks, is currently centered around the device and communication block, while the rest of the blocks will be considered in the future work. In addition, Section 2.2, we provide a thorough analysis of the QoE concept with respect to modeling, measuring, and predicting QoE. Finally, we review the state of the art in the area of QoE for IoT and provide a classification table (Table 2.2) to compare it with the proposed contribution in Section 1.5.2.

Chapter 3

Conclusion and Future Work

In this chapter we conclude the thesis in light of the presented research questions, in Section 1.3, and thesis contributions, in Section 1.5. In addition, we provide research directions for future work.

3.1 Thesis conclusions

This thesis has identified a number of research challenges with respect to extending the definition of QoE to cover IoT services. Direct research on QoE has been conducted for several years, but mainly in the area of multimedia services, where the service performance is evaluated from the end-user's perspective [16, 59, 60, 73]. With the emerging IoT, we assert that the presence of intelligent machines ultimately requires defining new relationships with the end-users and studying their impact on the QoE. Consider a human-machine interaction within an IoT service, where end-users can supervise, alter, or override an autonomous process. The decision making in an autonomous process is based on Machine-to-machine (M2M) communication and, currently, reasoning capabilities supported by Artificial Intelligence (AI) [30]. The state of the art QoE management models are verified with subjective, human tests, which neither captures or understands the implications of quality degradation caused by the intelligent machines. Thus, as a first research question in this thesis, we raise the need of extending the current QoE definition for covering emerging IoT services, which we address with the following contribution:

Contribution 1 - Defining Quality of IoT experience (QoIoT). Firstly, we distinguish the main actors that can contribute, and experience, to the overall IoT service quality. Besides evaluating the experience from the end-user, human, perception, we assert the need of evaluating the Quality of Machine Experience (QoME). As discussed by Hoffman and Novak [35], the concept of machine experience does not require smart objects to have conscious experience in order to have some type of experience. Instead, the machine experience is a representations of the M2M interactions and AI reasoning. Further, we discuss a real-life industrial IoT service that involve remotely-controlled mining vehicles, which we utilize as a basis of creating metrics to evaluate the QoME.

Secondly, we combine the proposed QoME with the legacy, end-user's, QoE in a form of QoIoT. The scope of QoIoT is focused on services that involve co-existence of humans and intelligent machines, such as the mentioned mining case-study. We assert that QoIoT has the goal of improving the human-business and business-machine relationships, by producing metrics that can make the end-users aware of the machines and vice versa.

Contribution 2. - Developing an architecture to model Quality of IoT experience. The second contribution answers a question that arises as a follow up to the first one, which is developing an architecture to model the QoIoT. The proposed QoIoT architecture consists of three steps, to model, measure, and predict QoIoT. Each of the steps have a pre-defined set of input metrics and processes that compute output metrics, specifically tailored for the mining case-study. Firstly, *model QoIoT* detects the main data-source, monitors the raw data produced in the service by classifying it within physical, network, application, and virtual layers. Secondly, *measure QoIoT* extracts the data analysis process from the perception layers and incorporates processes that evaluate the produced data and metrics from the perspective of Quality of Data, Network, and Context. The overall goal of this entity is to produce metrics that will evaluate the performance of the IoT service from the human and machine perspective, i.e. QoE and QoME. We identify metrics such as productivity, efficiency, and reliability as the most important in the mining case-study. Thirdly, *predict QoIoT* is envisioned to have a global view of the overall service by considering the metrics produced from the previous layers. Then, it takes on forecasting methods and attempt to predict the mentioned QoIoT metrics in the future.

Contribution 3 - Validating the proposed architecture for Quality of IoT experience. The scope of the validation is currently centered on the *measure QoIoT*, by evaluating metrics that describe the Quality of Network (QoN). We develop a, non-intrusive, Machine Learning (ML) models for estimating latency and a number of video quality assessment (VQA) metrics regarding IoT devices and video streaming, respectively, within the mining case-study. Driving tests were conducted in real-life mining environment, with aim to evaluate how different wireless conditions, such as coverage, network load, interference, and doppler, impact the latency and VQA. Regarding the feature engineering part, we utilize wireless access link characteristics, describing the cellular radio. Firstly, the achieved results suggest that it to possible to classify the IoT communication based on its sampling frequency to estimate the latency. Regarding 20, 50, and 100ms of sampling period, we have achieved 90, 85, and 82% coefficient of determination (R^2), respectively. Secondly, subjective and objective VQA metrics, such as MOS, PSNR, Frame delay, Frame Skips, and Blurriness were calculated on 10 pre-recorded videos, which were streamed during drive tests in total of 200 times. The achieved coefficient of determination (R^2) was 94, 89, 84, 85, and 82% when estimating MOS, PSNR, Frame Delay, Frame Skips, and Blurriness, respectively. The scope of the study also considered a root cause analysis on a estimated poor MOS value. By analyzing the utilized ML model, we show that it is possible to build a range of values for each feature that will alert when the video streaming is expected to have a MOS value below 3.5.

3.2 Future work

There are several directions for future research, in order to complete and advance the work presented in this thesis. The most immediate future work is related to the validation process of the proposed QoIoT architecture. We envision placing the architecture in each of the autonomous mining vehicles to evaluate the QoIoT in real-time, with the aim of improving the service performance. The validation process encompasses mostly experimental research within the proposed layers in the *measure QoIoT* entity. We envision the future research to validate the whole Measure QoIoT, with respect to the proposed layers, Quality of Network (QoN), Quality of Data (QoD), and Quality of Context (QoC). For instance, an interesting approach would be to study methods that can pinpoint at failures, anomalies, and other performance metrics regarding the hardware and software of the IoT devices, regarding QoD. Further, we consider a thorough understanding of the service requirements and signed SLAs within the mining case in order to produce a real-time calculation of the proposed QoIoT metrics - productivity, efficiency, and reliability. The calculation of the proposed QoIoT metrics would ultimately enable a validation of the final step within the QoIoT architecture - Predict QoIoT, where we envision forecasting the proposed metrics.

Regarding high-level view of the proposed QoIoT architecture, we suggest utilizing the technical practices and learnings from the mining case-study to deploy QoIoT in various other interdisciplinary domains, such as in health-care (e.g. remotely controlled surgery), smart manufacturing, smart city, and mitigating natural disaster services.

References

[1] 3GPP. LTE: Physical layer measurements. TS 36.214, release 10. 2013, accessed online (March 2020).

[2] 3GPP. 3GPP TS 23.501: System architecture for the 5G System, Release 16. 2019, accessed online (March 2020).

[3] Ibrahim Afolabi, Tarik Taleb, Konstantinos Samdanis, Adlen Ksentini, and Hannu Flinck. Network slicing and softwarization: A survey on principles, enabling technologies, and solutions. *IEEE Communications Surveys & Tutorials*, 20(3):2429–2453, 2018.

[4] Farhana Afroz, Ramprasad Subramanian, Roshanak Heidary, Kumbesan Sandrasegaran, and Solaimanitu2006qoe Ahmed. Sinr, rsrp, rssi and rsrq measurements in long term evolution networks. *International Journal of Wireless & Mobile Networks*, 2015.

[5] Mustafa Riza Akdeniz, Yuanpeng Liu, Mathew K Samimi, Shu Sun, Sundeep Rangan, Theodore S Rappaport, and Elza Erkip. Millimeter wave channel modeling and cellular capacity evaluation. *IEEE journal on selected areas in communications*, 32(6):1164–1179, 2014.

[6] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE communications surveys & tutorials*, 17(4):2347–2376, 2015.

[7] Mohammed Alreshoodi and John Woods. Survey on qoe\qos correlation models for multimedia services. *arXiv preprint arXiv:1306.0221*, 2013.

[8] Giuseppe Araniti, Claudia Campolo, Massimo Condoluci, Antonio Iera, and Antonella Molinaro. Lte for vehicular networking: a survey. *IEEE communications magazine*, 51(5):148–157, 2013.

[9] Álvaro Arcos-García, Juan A Alvarez-Garcia, and Luis M Soria-Morillo. Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods. *Neural Networks*, 99:158–165, 2018.

[10] João Ascenso, Hudson Cruz, and Patrícia Dias. Packet-header based no-reference quality metrics for h. 264/avc video transmission. In *2012 International Conference on Telecommunications and Multimedia (TEMU)*, pages 174–151. IEEE, 2012.

[11] Kevin Ashton. That 'internet of things' thing. *RFID journal*, 22(7):97–114, 2009.

[12] Arshdeep Bahga and Vijay Madisetti. *Internet of Things: A hands-on approach*. Vpt, 2014.

[13] Sabina Baraković and Lea Skorin-Kapov. Survey and challenges of qoe management issues in wireless networks. *Journal of Computer Networks and Communications*, 2013, 2013.

[14] John Boardman and Brian Sauser. System of systems-the meaning of of. In *2006 IEEE/SMC International Conference on System of Systems Engineering*, pages 6–pp. IEEE, 2006.

[15] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.

[16] Kjell Brunnström, Sergio Ariel Beker, Katrien De Moor, Ann Dooms, Sebastian Egger, Marie-Neige Garcia, Tobias Hossfeld, Satu Jumisko-Pyykkö, Christian Keimel, and Mohamed-Chaker Larabi. Qualinet white paper on definitions of quality of experience. 2013.

[17] Pedro Casas, Alessandro D'Alconzo, Florian Wamser, Michael Seufert, Bruno Gardlo, Anika Schwind, Phuoc Tran-Gia, and Raimund Schatz. Predicting qoe in cellular networks using machine learning and in-smartphone measurements. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2017.

[18] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.

[19] Yanjiao Chen, Kaishun Wu, and Qian Zhang. From qos to qoe: A tutorial on video quality assessment. *IEEE Communications Surveys & Tutorials*, 17(2):1126–1165, 2014.

[20] Li Da Xu, Wu He, and Shancang Li. Internet of things in industries: A survey. *IEEE Transactions on industrial informatics*, 10(4):2233–2243, 2014.

[21] Rahul Dagar, Subhranil Som, and Sunil Kumar Khatri. Smart farming–iot in agriculture. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 1052–1056. IEEE, 2018.

[22] Jerker Delsing. *Iot automation: Arrowhead framework*. CRC Press, 2017.

[23] Konstantinos Douzis, Stelios Sotiriadis, Euripides GM Petrakis, and Cristiana Amza. Modular and generic iot management on the cloud. *Future Generation Computer Systems*, 78:369–378, 2018.

[24] Mica Endsley. Toward a theory of situation awareness in dynamic systems. In *Situational Awareness*, pages 9–42. Routledge, 2017.

[25] Ericsson. Ericsson mobility report, November 2019. Whitepaper, 2019, accessed online (March 2020).

[26] Hazem M Fahmy, Mohamed A Abd El Ghany, and Gerd Baumann. Vehicle risk assessment and control for lane-keeping and collision avoidance at low-speed and high-speed scenarios. *IEEE Transactions on Vehicular Technology*, 67(6):4806–4818, 2018.

[27] Markus Fiedler, Tobias Hossfeld, and Phuoc Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *IEEE Network*, 24(2):36–41, 2010.

[28] Alessandro Floris and Luigi Atzori. Managing the quality of experience in the multimedia internet of things: A layered-based approach. *Sensors*, 16(12):2057, 2016.

[29] Ammar Gharaibeh, Mohammad A Salahuddin, Sayed Jahed Hussini, Abdallah Khreishah, Issa Khalil, Mohsen Guizani, and Ala Al-Fuqaha. Smart cities: A survey on data management, security, and enabling technologies. *IEEE Communications Surveys & Tutorials*, 19(4):2456–2501, 2017.

[30] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[31] Michael A Goodrich and Alan C Schultz. Human–robot interaction: a survey. *Foundations and Trends® in Human–Computer Interaction*, 1(3):203–275, 2008.

[32] Luca Greco, Pierluigi Ritrovato, Thanassis Tiropanis, and Fatos Xhafa. Iot and semantic web technologies for event detection in natural disasters. *Concurrency and Computation: Practice and Experience*, 30(21):e4789, 2018.

[33] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7):1645–1660, 2013.

[34] Tobias Hobfeld, Raimund Schatz, Martin Varela, and Christian Timmerer. Challenges of qoe management for cloud applications. *IEEE Communications Magazine*, 50(4):28–36, 2012.

[35] Donna L Hoffman and Thomas P Novak. Consumer and object experience in the internet of things: An assemblage theory approach. *Journal of Consumer Research*, 44(6):1178–1204, 2018.

[36] Hilary J Holz, Anne Applin, Bruria Haberman, Donald Joyce, Helen Purchase, and Catherine Reed. Research methods in computing: what are they, and how should we teach them? *ACM SIGCSE Bulletin*, 38(4):96–114, 2006.

[37] Xiaohong Huang, Kun Xie, Supeng Leng, Tingting Yuan, and Maode Ma. Improving quality of experience in multimedia internet of things leveraging machine learning on big data. *Future Generation Computer Systems*, 86:1413–1423, 2018.

[38] ITU. ITU-T P.800 Recommendation: Methods for subjective determination of transmission quality. 1998, accessed online (March 2020).

[39] ITU. ITU-T P.10/G.100 Amendment 1: Definition of QoE. 2006, accessed online (March 2020).

[40] ITU. ITU-T J.247 Recommendation: Objective perceptual multimedia video quality measurement in the presence of a full reference. 2008, accessed online (March 2020).

[41] ITU. ITU-T Recommendations Y.4000/Y.2060: Overview of the internet of things. 2012, accessed online (March 2020).

[42] ITU. ITU-T G.107 Recommendation: The e-model: a computational model for use in transmission planning. 2015, accessed online (March 2020).

[43] ITU. ITU-T P.1203 Recommendation: Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport. 2017, accessed online (March 2020).

[44] ITU. ITU-T P.863 Recommendation: Perceptual objective listening quality prediction. 2018, accessed online (March 2020).

[45] ITU. ITU-T Y.1540: internet protocol data communication service - IP packet transfer and availability performance parameters. 2019, accessed online (March 2020).

[46] Sabina Jeschke, Christian Brecher, Tobias Meisen, Denis Özdemir, and Tim Eschert. Industrial internet of things and cyber manufacturing systems. In *Industrial Internet of Things*, pages 3–19. Springer, 2017.

[47] Ingemar Johansson. Self-clocked rate adaptation for conversational video in lte. In *Proceedings of the 2014 ACM SIGCOMM workshop on Capacity sharing workshop*, pages 51–56. ACM, 2014.

[48] Amulya Karaadi, Lingfen Sun, and Is-Haka Mkwawa. Multimedia communications in internet of things qot or qoe? In *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pages 23–29. IEEE, 2017.

[49] Chakravanti Rajagopalachari Kothari. *Research methodology: Methods and techniques*. New Age International, 2004.

[50] Li Li, Yi-Lun Lin, Nan-Ning Zheng, Fei-Yue Wang, Yuehu Liu, Dongpu Cao, Kunfeng Wang, and Wu-Ling Huang. Artificial intelligence test: a case study of intelligent vehicles. *Artificial Intelligence Review*, 50(3):441–465, 2018.

[51] Olof Liberg, Marten Sundberg, Eric Wang, Johan Bergman, and Joachim Sachs. *Cellular Internet of things: technologies, standards, and performance*. Academic Press, 2017.

[52] Jie Lin, Wei Yu, Nan Zhang, Xinyu Yang, Hanlin Zhang, and Wei Zhao. A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. *IEEE Internet of Things Journal*, 4(5):1125–1142, 2017.

[53] Yi Liu, Chao Yang, Li Jiang, Shengli Xie, and Yan Zhang. Intelligent edge computing for iot-based energy management in smart cities. *IEEE Network*, 33(2):111–117, 2019.

[54] Kais Mekki, Eddy Bajic, Frederic Chaxel, and Fernand Meyer. A comparative study of lpwan technologies for large-scale iot deployment. *ICT express*, 5(1):1–7, 2019.

[55] Dimitar Minovski, Christer Åhlund, and Karan Mitra. Modeling quality of iot experience in autonomous vehicles. *IEEE Internet of Things Journal*, pages 1–1, 2020.

[56] Dimitar Minovski, Christer Åhlund, Karan Mitra, and Per Johansson. Analysis and estimation of video qoe in wireless cellular networks using machine learning. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2019.

[57] Dimitar Minovski, Christer Åhlund, Karan Mitra, and Roman Zhohov. Quality of experience for the internet of things. *IEEE IT Professional Magazine*, 2020.

[58] Marvin Minsky. Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1):8–30, 1961.

[59] Karan Mitra, Arkady Zaslavsky, and Christer Åhlund. Context-aware qoe modelling, measurement, and prediction in mobile computing systems. *IEEE Transactions on Mobile Computing*, 14(5):920–936, 2013.

[60] Sebastian Möller and Alexander Raake. *Quality of experience: advanced concepts, applications and methods*. Springer, 2014.

[61] Subhas Chandra Mukhopadhyay and Nagender K Suryadevara. Internet of things: Challenges and opportunities. In *Internet of Things*, pages 1–17. Springer, 2014.

[62] Junghoon Oh and Hoyoung Song. Study on the effect of lte on the coexistence of nb-iot. In *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 610–612. IEEE, 2018.

[63] Irena Orsolic, Mirko Suznjevic, and Lea Skorin-Kapov. Youtube qoe estimation from encrypted traffic: Comparison of test methodologies and machine learning based models. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2018.

[64] Verdes Pedras, Marco Sousa, Pedro Vieira, Maria Paula Queluz, and Antonio Rodrigues. A no-reference user centric qoe model for voice and web browsing based on 3g/4g radio measurements. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2018.

[65] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):1175–1191, 2001.

[66] Petar Popovski, Kasper Fløe Trillingsgaard, Osvaldo Simeone, and Giuseppe Durisi. 5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view. *IEEE Access*, 6:55765–55779, 2018.

[67] Rapeepat Ratasuk, Benny Vejlgaard, Nitin Mangalvedhe, and Amitava Ghosh. Nb-iot system for m2m communication. In *2016 IEEE wireless communications and networking conference*, pages 1–5. IEEE, 2016.

[68] Werner Robitza, Arslan Ahmad, Peter A Kara, Luigi Atzori, Maria G Martini, Alexander Raake, and Lingfen Sun. Challenges of future multimedia qoe monitoring for internet service providers. *Multimedia Tools and Applications*, 76(21):22243–22266, 2017.

[69] N Shabana and G Velmathi. Advanced tele-surgery with iot approach. In *Intelligent Embedded Systems*, pages 17–24. Springer, 2018.

[70] Shree Krishna Sharma and Xianbin Wang. Towards massive machine type communications in ultra-dense cellular iot networks: Current issues and machine learning-assisted solutions. *IEEE Communications Surveys & Tutorials*, 2019.

[71] Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5):637–646, 2016.

[72] Dong-Hee Shin. Conceptualizing and measuring quality of experience of the internet of things: Exploring how quality is perceived by users. *Information & Management*, 54(8):998–1011, 2017.

[73] Lea Skorin-Kapov, Martín Varela, Tobias Hoßfeld, and Kuan-Ta Chen. A survey of emerging concepts and challenges for qoe management of multimedia services.

ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 14(2s):1–29, 2018.

[74] Stanley Solow. Wireless remote controlled electronic system for a vehicle, October 16 2018. US Patent 10,099,635.

[75] Filip Stojanovski. Churn prediction using sequential activity patterns in an on-demand music streaming service, 2017.

[76] Muhammad Suryanegara, Dimas Agung Prasetyo, Fery Andriyanto, and Nur Hayati. A 5-step framework for measuring the quality of experience (qoe) of internet of things (iot) services. *IEEE Access*, 7:175779–175792, 2019.

[77] Rob Van der Mei, Hans Van den Berg, Ivan Ganchev, Kurt Tutschku, Philipp Leitner, Pasi Lassila, Wojciech Burakowski, Fidel Liberal, Åke Arvidsson, Tobias Hoβfeld, et al. State of the art and research challenges in the area of autonomous control for a reliable internet of services. In *Autonomous Control for a Reliable Internet of Services*, pages 1–22. Springer, Cham, 2018.

[78] Michael Welzl, Safiqul Islam, and Stein Gjessing. Coupled congestion control for rtp media. *Transport*, 6:1, 2013.

[79] Yu-Ting Wen, Pei-Wen Yeh, Tzu-Hao Tsai, Wen-Chih Peng, and Hong-Han Shuai. Customer purchase behavior prediction from payment datasets. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 628–636. ACM, 2018.

[80] Qihui Wu, Guoru Ding, Yuhua Xu, Shuo Feng, Zhiyong Du, Jinlong Wang, and Keping Long. Cognitive internet of things: a new paradigm beyond connection. *IEEE Internet of Things Journal*, 1(2):129–143, 2014.

[81] Junyong You, Ulrich Reiter, Miska M Hannuksela, Moncef Gabbouj, and Andrew Perkis. Perceptual-based quality assessment for audio–visual services: A survey. *Signal Processing: Image Communication*, 25(7):482–501, 2010.

[82] Chaoqun Yue, Ruofan Jin, Kyoungwon Suh, Yanyuan Qin, Bing Wang, and Wei Wei. Linkforecast: Cellular link bandwidth prediction in lte networks. *IEEE Transactions on Mobile Computing*, 17(7):1582–1594, 2017.

[83] Lin Zhang, Ying-Chang Liang, and Ming Xiao. Spectrum sharing for internet of things: A survey. *IEEE Wireless Communications*, 26(3):132–139, 2018.

[84] Roman Zhohov, Dimitar Minovski, Per Johansson, and Karl Andersson. Real-time performance evaluation of lte for iiot. In *2018 IEEE 43rd Conference on Local Computer Networks (LCN)*, pages 623–631. IEEE, 2018.

Part II: Included Papers

Paper A

Defining Quality of Experience for the Internet of Things

Authors:
Dimitar Minovski, Chrsiter Åhlund, Karan Mitra, Roman Zhohov

Reformatted version of paper originally published in:
IEEE IT Professional Magazine, 2020.

Defining Quality of Experience for the Internet of Things

Dimitar Minovski, Chrsiter Åhlund, Karan Mitra, Roman Zhohov

Abstract

The Internet of Things (IoT) brings a set of unique and complex challenges to the field of Quality of Experience (QoE) evaluation. The state-of-the-art research in QoE mainly targets multimedia services, such as voice, video, and the Web, to determine quality perceived by end-users. Therein, main evaluation metrics involve subjective and objective human factors and network quality factors. Emerging IoT may also include intelligent machines within services such as health-care, logistics, and manufacturing. The integration of new technologies such as Machine-to-Machine communications and artificial intelligence within IoT services may lead to service quality degradation caused by machines. In this article, we argue that evaluating QoE in the IoT services should also involve novel metrics for measuring the performance of the machines alongside metrics for end-users' QoE. This article extends the legacy QoE definition in the area of IoT and defines conceptual metrics for evaluating QoE using an industrial IoT case-study.

1 Introduction

IoT are physical objects equipped with Information and Communication Technologies (ICT) connecting to the Internet [20]. Figure 1 shows examples of emerging IoT services such as health-care, logistics, industrial production, and robotics. The unique nature of these services leads to a range of different quality requirements that are difficult to tackle [17]. Consider autonomous vehicles as an example, where the focus is on fast and accurate processing of sensory data to avoid collisions; while IoT-enabled remote surgery service imposes stringent latency requirements. Quality provisioning in these services is not only limited to life-threatening situations but also to reduce the risk of causing significant business losses and environmental damage.

From a stakeholder perspective (e.g., service providers and owners), service quality evaluation is an indispensable business goal for harmonizing the end-users' experiences (QoE), agreed upon through a Service Level Agreement (SLA) [19]. The area of QoE deals with the understanding of end-users' perception of the services offered by the stakeholder [11,16]. However, the conventional QoE techniques [16] mainly target multimedia communications and do not consider the performance and impact of Machine-to-Machine (M2M) communication and autonomous processes on QoE. The emerging IoT services on Figure 1 may involve humans and intelligent machines, which can interact with each

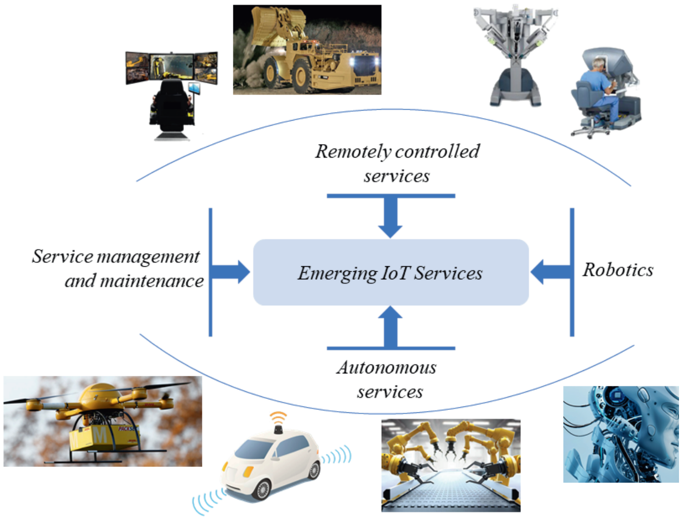


Figure 1: The emerging IoT use-cases.

other through Human-Computer Interaction (HCI) interfaces. Herein, the HCI interaction becomes increasingly complicated as the end-user can provide inputs and alter the decisions by the machines. Moreover, autonomous processes typically consist of multiple predictive models [17] which take decisions based on various data processing and M2M communications, where any hardware and software part can be subject to failure and potentially affect the QoE. Thus, this article argues for a paradigm shift in QoE area to cover IoT services where, besides end-users, the quality and performance evaluation should also target intelligent machines, as well as their implications on the end-users and businesses. As a consequence, the main contributions of this article are:

- Identifying the main challenges of assessing QoE for IoT by adding the complexity of autonomous processes and machine experience to the current understandings of QoE.
- Defining the term Quality of IoT-experience (QoIoT) from the perspective of both humans and machines to understand the human-business and machine-business relationships. From the stakeholder perspective, QoIoT is beneficial to determine the overall service performance, as well as beneficial for the end-users and machines to understand their impact on service performance.

2 Background on QoE for IoT

Traditional Quality of Service (QoS) parameters such as delay, jitter, packet loss, and bandwidth are used to evaluate network quality objectively [11]. However, QoS narrowly captures the end user's perception of the service and may not reflect their overall service perception or experience. To highlight the end-user's point-of-view, the concept of QoE was proposed as: *"The overall acceptability of an application or service, as perceived subjectively by the end-user"* [12]. Since then, the research on QoE has focused from mapping QoS metrics to users' QoE [7]; to understanding and defining factors influencing QoE [11, 14]; and to building QoE management models [16]. The research on QoE management is addressed from two distinct, but often complementary perspectives, *application* and *network* provider [3, 16]. In particular, QoE-driven application management has focused mainly on end-users' perception of multimedia and Web-services [16]. Network QoE management focuses on the way end-users perceive network related processes. For instance, it measures QoS-related parameters to describe the performance of the cellular networks, base-stations, and communication protocols [3]. The advancements in network technologies such as Software Defined Networks (SDN) and Network Function Virtualization (NFV) bring additional complexities when measuring both application and network QoE. For instance, network resource allocation, storage, and computation can now be virtually shared among end-users [1]; this dramatically reduces the operational costs but complicates the overall performance benchmarking as there is no standardized way for SDN and NFV orchestration yet [4].

Understanding QoE within IoT is still in its infancy; these areas are discussed widely; however, in isolation [9,12]. There are very few studies that directly target QoE evaluation of IoT services [8, 20]; and to the best of our knowledge there is a complete absence of studies regarding QoE in Industrial IoT. Wu et al. [20] present one of the earliest works discussing QoE for IoT. They propose the concept of Cognitive IoT as a new network paradigm where physical and virtual objects are interconnected and perform as agents. In addition, Wu et al. [20] identified layers of data in an IoT service to be considered for QoE evaluation. Namely, they distinguish a data layer from information layer, where the former one aims to evaluate the quality of the sensed data, while the latter one deals with contextual meaning of the data. Floris and Atzori [8] take a step further in developing a concept of layered-QoE framework to evaluate the quality of a multimedia IoT service. Therein, a key method for Quality of Data (QoD) is proposed, where the accuracy of the produced IoT data is benchmarked. Although the authors do not provide detailed steps on how to measure QoD, they use an abstracted version of it in combination with QoS to map them to subjective QoE score. However, the authors also provide no indications on how their model could cover other IoT services except multimedia, as shown in Figure 1. Our study builds upon the works of [8, 20] by significantly expanding the discussion in the direction of emerging IoT services and the way end-users interact with them.

The expansion of IoT in automation and intelligent decision-making brings a set of challenges which are not discussed in current state of the art in QoE for IoT. Figure 2 shows an example of an emerging IoT service that integrates a combination of functionalities provided by other services, with input and output data coming from both humans and machines [9]. Such example can lead to complex service chains with possibly even hundreds of services offered by different third parties, each with their business incentives. Therein, a single decision by an autonomous process may be taken as a result of multiple data sources that are dependent on one another. Consider the risk assessment of a safety alarm service in an underground mine as example, which consists of multiple sensors to detect smoke, humidity, and temperature. An autonomous process triggers the alarm in case of an accident, which is followed by a series of chained events. Then, a process locates the miners and computes their most optimal rescue route, using the data coming from different sets of sensor sources. Moreover, the service must notify the miners for their own rescue route in a timely manner and contact additional external services to deal with the emergency situations, such as the fire department in case of a fire. Potential sources of failures in such systems are multi-dimensional, arising from: (1) IoT hardware and software, with respect to data quality [20]; (2) Network, related to delays and packet losses [3]; and (3) Virtual computing, with respect to data processing, caching, and storage [11]. In current practice, the quality of such complex chained events is controlled on an ad-hoc basis, while the consequences of failures are not well understood [17]. The same study [17] also identifies current lack of understanding on QoE measurement and monitoring regarding multiple chained events and autonomous services. Thus, the aim of the next section is to identify the components and data-sources of the emerging IoT services.

3 Quality of Machine Experience (QoME)

Artificial Intelligence (AI) is seen as a key enabler for autonomous processes [9], which can also be part of complex service chains as shown on Figure 2. AI is defined as: *"the science of making machines capable of performing tasks that would require intelligence if done by humans"* [10]. The research on AI goes as far as defining machine emotions and human-robot interaction [13]. Thus, one may envision that AI-enabled machines can, just like humans, perceive experience, which arguably is different compared to the humans (e.g., QoE). The AI runs as a software on a controller (Figure 2), which takes decisions based on data-driven approaches [10]. Moreover, decision making requires reasoning, which entails complicated processes for self-optimization and self-configuration [9]. For instance, self-optimizing algorithms anticipate critical incidents and other occurrences of the operating history to optimize the solution behavior. Thus, we define Quality of Machine Experience (QoME) as: *"Objective metrics that measures the quality and performance of intelligent machines and their decisions"*. Herein, the first challenge is to identify the main components and data-sources regarding intelligent machines; Secondly, the objective metrics should describe the current state of the system in a humanly understandable way. The creation of such objective metrics is essentially the scope of this article, proposed with the QoIoT definition in the next section.

The data-sources within the AI are typically classified into three components, namely (1) physical, (2) network, and (3) logical, explained as [9, 12]:

1. Physical components are related to the data generated by the hardware and its software withing an service. The state of the physical components is measured through behavior monitoring, fault tolerance, and anomaly detection in the generated data [17]. To do so, researchers study statistical methods and unsupervised Machine Learning (ML) techniques to find relationships and clusters in the generated data [9].

2. Network components are related to the research on QoS with main focus on mea-

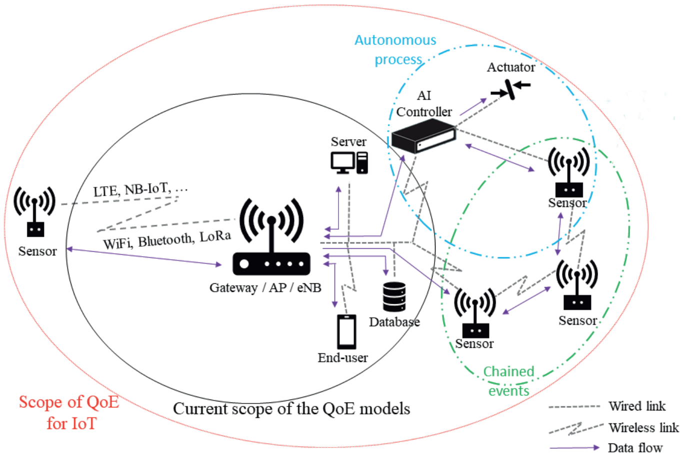


Figure 2: QoE in an end-to-end IoT service.

suring the way physical devices and end-users communicate within the service and with the outside world [11].

3. Logical components are related to the software, or controller, that typically has various IoT hardware on disposal to execute the following processes [9]: (a) Independent decision making based on contextual conditions; (b) Prediction of states, events and conditions; (c) Event-based M2M interaction patterns; (d) Autonomous creation of knowledge and utilizing it for self-optimization; (e) Self-reliant reaction due to unplanned events. Those processes can be quantified using objective data generated by the software and its physical devices. However, current literature does not understand the way such autonomous services impacts the overall QoS and QoE [17].

4 Challenges regarding QoE for IoT

The following list expose the limitations of the current state of the art methods for monitoring and evaluating quality of an IoT service:

1. How does various physical, network, and logical components affect each other in a complex service chains?
2. How to perform a root-cause analysis in service chains, containing intelligent machines?
3. How can an end-user alter or affect the intelligent machines?
4. How to evaluate the quality of intelligent machines and troubleshoot their impact on the service performance?
5. With what type of objective metrics for quality evaluation and control to feed the self-optimizing methods in an autonomous service?

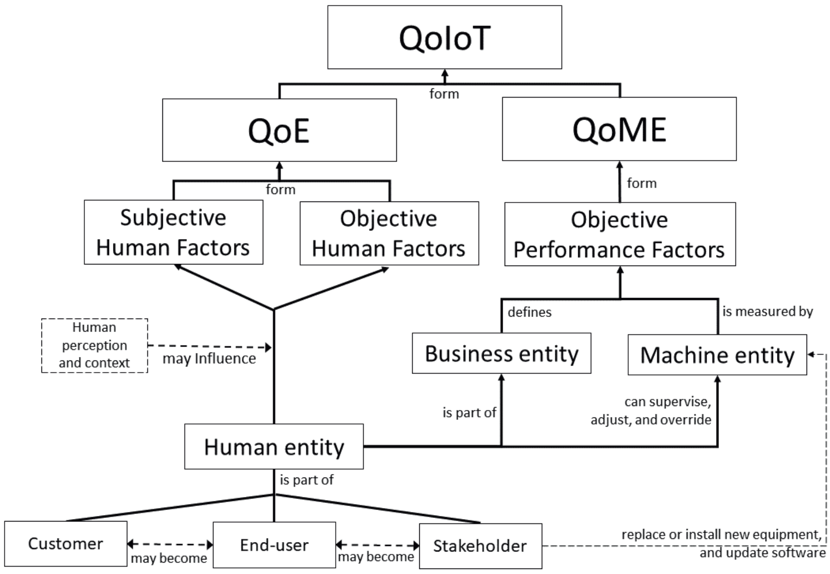


Figure 3: Defining QoIoT.

5 Defining Quality of IoT-experience (QoIoT)

The legacy models for QoE management are user-centric, quantifying the experience from the human entity through subjective and objective tests [3, 11, 16]. The human entity is composed of end-users, which in [11] are seen as a customers from a business perspective. In this regard, the concept of Quality of Business (QoB) [18] was developed, which looks at the impact of QoE and QoS to business operations, defining relationship such as business to customer and business to business. The end-user interacts with the service (Figure 2), even though the service may consist of autonomous processes and complex service chains. Consider an autonomous vehicle, where an end-user can supervise, adjust the settings, or override a decision by made by the AI [6]. However, before taking any decision to alter the autonomy, the end-users need knowledge of the current system's state, referred as Situation Awareness (SA) [5]. Going beyond the QoE approaches, the decisions by the intelligent machines will impact the business entity as well, sometimes without human intervention [17]. In that direction, we propose an extension to the quality evaluation paradigm to cover emerging IoT services by defining Quality of IoT-experience (QoIoT), illustrated in Figure 3. Essentially QoIoT is a combination of metrics that are measured from two different perspectives: (1) QoE, by extracting the legacy user-centric management models [3, 11, 16]; and, (2) QoME, by understanding machine's behavior and translating the overall machine performance to business-oriented metrics. QoIoT is represented as:

$$QoIoT \in \{QoE, QoME\} \tag{1}$$

Correspondingly, the newly formed definition of QoIoT states as follows: *"Quality of IoT-experience (QoIoT) is a metric that aggregates the delivered quality of an IoT service from the perspective of humans and machines. The end-users' experience is evaluated with*

Components in an IoT service	*Key Performance Indicators*	*Classification*
Physical	*Hardware type (e.g. sensors, actuators, cameras, lidar), CPU, OS, raw data, labels, memory, display resolution, data-type format, sampling frequency, sensing characteristics, energy consumption, cryptography*	*QoS*
Network	*Jitter, packet-loss, bandwidth, delay, buffer size, Received-Signal-Strength Indicator (RSSI), Reference Signal Received Quality (RSRQ), Signal-to-Noise Ratio (SNR), etc.*	
Application and service specific	*Application type (e.g. video, voice, IoT, web), Application metrics (e.g. bit-error rate, encoding, bitrate, content genre); Subjective human scores (e.g. MOS, mood, enjoyment, emotions, easy-of-use, end-user's characteristics and background), Objective human scores (e.g. HVS)*	*QoE*
Logical system components	*Benchmarking individual data-sources (e.g. anomaly and novelty detection, context-aware pattern recognition), benchmarking predictive models (e.g. condition monitoring, diagnostics, predictive maintenance), benchmarking the impact of autonomous processes on overall service performance (e.g. productivity, safety, reliability)*	*QoME* *QoIoT*

Figure 4: Synergies between QoS, QoE, QoME, and QoIoT.

subjective and objective human factors, while Quality of Machine Experience (QoME) translates the overall machine performance to business-oriented metrics through objective performance indicators."

5.1 Human-business relationship

QoIoT, by definition, extends the traditional user-centric QoE approaches and therefore extracts the Human Entity, which includes stakeholders, end-users, and customers [12], as shown in Figure 3. The current state of the art also defines the relationship between human and business [19], mainly by utilizing QoS and QoE to build SLAs. Hence, the Business Entity in the current form understands the customer needs and possess abilities to alter the business operations through defining business goals, SLAs, and build the IoT infrastructure. Figure 4 summarizes the synergies among QoIoT with the current state of the art, classified by the identified components of an IoT service [9, 12, 17]. The first criterion in Figure 4 is QoS, which strictly describes the network quality with objectively measured metrics [11]. Majority of the QoE management models in [16] utilize the QoS metrics, however, their relationship is not linear; this is due to human cognitive and behavior factors, as well as various contextual factors [11]. That is, QoE metrics consider factors from the application and network metrics and ultimately maps them to end-users' subjective scores. Such subjective knowledge provisions the end-user experience, but do not calibrate the overall machine performance or enable detailed root-causes of the service performance. For instance, policymakers for the autonomous vehicles are struggling in defining the responsible actors in cases of accidents [6], where subjective scores do not help in determining whether it is a human mistake or point out at a specific internal hardware or software error.

5.2 Business-machine relationship

Endsley [5] defines the term SA in dynamic systems; it means to observe the current state of the system, detect patterns, and create objective knowledge necessary for service control which has direct implications on the overall service performance. Therefore, in complex service chains composed of end-users and autonomous services, it is essential to create QoIoT knowledge understandable for both entities, humans, and machines. On the one hand, measuring QoIoT can leverage the Human entity regarding: (1) Understanding the impact of the machines on the business and user-centric QoE; (2) Replace or install new equipment, and upgrade software; (3) Semi-controlling the machines; and (4) Adjust the workflow of the machines. On the other hand, the QoIoT knowledge can leverage the Machine entity as well, regarding: (1) Understanding the impact of individual data-sources and the logical components of the intelligent machines on the overall service performance; and (2) Using the QoIoT metrics as benchmarks for optimizing the autonomous operations. In this regard, QoIoT considers measuring objective performance factors (Figure 4). Figure 4 lists the factors that concern various components for an IoT service; these factors measure how well an IoT service makes progress in achieving a particular objective such as productivity and safety. Each of those factors may be rather general, such as Overall Equipment Efficiency (OEE) [15] or industry-specific, such as revenue/hour. The book by Samsonowa [15] describes several examples of performance factors in Information Technology industry. A study on industrial manufacturing systems [9] gives a hint that quality of autonomous services can certainly be measured through business metrics, independent of any network and application technologies. Hence, the way performance factors are defined will depend on the business goals and service requirements. The next sub-section provides an example of forming such performance factors around a case-study.

6 Industrial IoT Case Study: Remotely controlled semi-autonomous underground vehicles

Remotely controlled mining vehicles are considered as a case study to apply the proposed QoIoT concept and define performance factors for quality monitoring. The motivation is to automate the operations in mine, such as blasting and extracting ore and remove the expert-driver from such hazardous underground environment. This would also fully utilize the production hours of the mine, especially during shift changes and blast hours, as the mine workers typically wait few hours after a blast to access the mine due to polluted air [2]. As an enabler for such case-study, an IoT setup deploys a variety of sensors, actuators, and cameras, streaming data in uplink over the cellular network to a remote-control station. As an expert-driver controls the machines from a distance, data is being sent in downlink as well, from joysticks, buttons, and pedals. The variety of traffic sources impose complex quality requirements for the network, as the traffic consists of critical and non-critical information, substantial and light-weighted data.

Besides remote-control from a distance, the expert-driver can also supervise semi-autonomous features and let the vehicle execute task autonomously. For instance, the vehicles can drive in-and-out of the mine independently, evacuate in case of emergency, automatically place and perform the blasting, and assist in loading the ore. Being semi-autonomous means that particular software is deployed on the vehicles that take autonomous decisions in controlling the vehicle. As the operations are orchestrated purely by centralized software, the run-time of the system must integrate complex processes such as decision-making, self-configuration, self-optimization, and self-healing [9].

6.1 Defining industry specific QoME factors around a case-study

The mining industry has pre-defined metrics which are constantly measured, based on which the business operations are optimized. Andersson [2] identified industrial revenue, safety, and productivity as key performance indicators within mining. While other businesses generate revenue from efficient and reliable service, in mining it is the productivity as the main factor determining the revenue. The productivity as a metric is a function of the productive capacity meeting its intended end-results, such as planned production levels, safety, efficiency, reliability, and direct economic costs [2]. Estimating the productivity in the mining case-study necessitates behavior monitoring of the expert drivers, machines, and network. Such role separation enables the possibility for root-cause analysis when the productivity levels are not at the desired level. A real-time productivity metric can enhance the expert-driver's SA during remote-control or supervision of the autonomous processes, regarding enabling/overriding a self-drive or adjusting the settings of the AI. Having the expert-driver intervene during self-driving is beneficial for the AI system itself, as the human behavior and the measured productivity can be used to label edge cases in improving the AI.

The productivity, in the mining case-study, may be defined with the following objective performance factors:

$$\begin{aligned} Productivity \leftarrow Production\ level\ \vee\ Reliability \\ \vee\ Efficiency \end{aligned} \tag{2}$$

The following list explores the ways to monitor the objective performance factors:

1. *Production level:* is defined as loading the maximum possible scoop of ore in the least amount of time, rendered as a tons/hour [2]. Herein, the essential metrics need to be identified, which lead to the desired tons/hour levels. For instance, weight, vibrations, trajectory, and pressure on the filling bucket can measure scooped tons/hour. A real-time ML analysis on such metrics, combined with internal vehicle's data, can compute the most suitable vehicle's acceleration while excavating

the ore [2]. In similar way can be computed the most optimal technique to load as much ore as possible per scoop, depending on the rock structure, directly increasing the tons/hour metric. In cases when an expert-driver is remotely-controlling the vehicle, real-time analysis on the network QoE, as well as video streaming QoE [16], can boost the overall productivity.

2. *Availability:* refers to the direct costs of repairing or replacing a broken machine in the mining case-study. However, the maintenance costs also include production losses during the downtime period of the broken machine, directly impacting productivity. A couple of machine availability threads are discussed in [2], such as wheel slip, collision detection and avoidance, and traction control. The same study also proposes ways to monitor the potential threads with the deployed sensors on the machines.

3. *Efficiency:* is associated with the success rate of correct operation, mean execution time, rate of operating errors, and complexity of the currently executing task. Fuel efficiency is one obvious metric in the context of the mining case-study, monitored and measured with a combination of sensors. For instance, the trajectory mappings of the underground mine with the positioning of the other machines or workers in real-time might show the most efficient driving path [2]. The same study also identifies the most important metrics for monitoring the efficiency in underground machines, such as bucket payload (tons), speed (km/h), transport performance (ton*km/h), and machine-utilization (%).

4. *Reliability:* refers to estimating the risk of failure in parts of the IoT service. An indication of how to measure the reliability of the case-study is given in [2], with metrics such as risk-of-failure, mean-time-to-failure, time-between-failure and reconfiguration, and mean-time-to-repair. For instance, to discover a percentage of risk-of-failure, one must monitor the overall machine health-status, in combination with the network analysis. Herein, thorough data analysis is necessary by utilizing methods for anomaly detection and fault tolerance, typically involving un-supervised machine learning techniques and data clustering [17]. An outcome can suggest when is the right time to replace hardware or upgrade software. Moreover, when the expert-driver remotely controls the vehicle, the system can indicate incoming blind radio coverage spots (where the driver loses connection with the machine) by observing the acceleration and position metrics from the machine.

7 Conclusion and future work

In this article, we identify a clear lack of definition for QoE in IoT. We envision the future emerging IoT services to be orchestrated purely from software, with and without human intervention. As the current state of the art on QoE is user-centric, the present QoE management models are quantified based on subjective tests, which neither captures or understands the implications of quality degradation caused by the machines. Thus, we assert the need for defining machine experience, where we aim to define a common ground for discussion on factors that influence the quality of autonomous services, regardless of the presence of the human. Within this context, it is the goal of this article is to delineate Quality of IoT-experience and to ignite discussions within the IoT and QoE community. As a future work, we aim to define and test a novel methodology for QoE assessment and prediction in IoT.

8 Acknowledgement

We thank Niclas Ögren and Per Johansson from InfoVista AB, the reviewers, and the editors for significantly improving this article.

References

[1] Naser Al-Falahy and Omar Y Alani. Technologies for 5g networks: Challenges and opportunities. *IT Professional*, 19(1):12–20, 2017.

[2] Ulf Andersson. *Automation and traction control of articulated vehicles*. PhD thesis, Luleå tekniska universitet, 2013.

[3] Sabina Baraković and Lea Skorin-Kapov. Survey and challenges of qoe management issues in wireless networks. *Journal of Computer Networks and Communications*, 2013, 2013.

[4] Mingzhe Chen, Walid Saad, and Changchuan Yin. Virtual reality over wireless networks: Quality-of-service model and learning-based resource management. *IEEE Transactions on Communications*, 66(11):5621–5635, 2018.

[5] Mica Endsley. Toward a theory of situation awareness in dynamic systems. In *Situational Awareness*, pages 9–42. Routledge, 2017.

[6] Daniel J Fagnant and Kara Kockelman. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, 77:167–181, 2015.

[7] Markus Fiedler, Tobias Hossfeld, and Phuoc Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *IEEE Network*, 24(2):36–41, 2010.

[8] Alessandro Floris and Luigi Atzori. Managing the quality of experience in the multimedia internet of things: A layered-based approach. *Sensors*, 16(12):2057, 2016.

[9] Sabina Jeschke, Christian Brecher, Tobias Meisen, Denis Özdemir, and Tim Eschert. Industrial internet of things and cyber manufacturing systems. In *Industrial Internet of Things*, pages 3–19. Springer, 2017.

[10] Li Li, Yi-Lun Lin, Nan-Ning Zheng, Fei-Yue Wang, Yuehu Liu, Dongpu Cao, Kunfeng Wang, and Wu-Ling Huang. Artificial intelligence test: a case study of intelligent vehicles. *Artificial Intelligence Review*, 50(3):441–465, 2018.

[11] Karan Mitra, Arkady Zaslavsky, and Christer Åhlund. Context-aware qoe modelling, measurement, and prediction in mobile computing systems. *IEEE Transactions on Mobile Computing*, 14(5):920–936, 2013.

[12] Sebastian Möller and Alexander Raake. *Quality of experience: advanced concepts, applications and methods*. Springer, 2014.

[13] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):1175–1191, 2001.

[14] Ulrich Reiter, Kjell Brunnström, Katrien De Moor, Mohamed-Chaker Larabi, Manuela Pereira, Antonio Pinheiro, Junyong You, and Andrej Zgank. Factors influencing quality of experience. In *Quality of experience*, pages 55–72. Springer, 2014.

[15] Tatjana Samsonowa. *Industrial research performance management: Key performance indicators in the ICT industry*. Springer Science & Business Media, 2011.

[16] Lea Skorin-Kapov, Martín Varela, Tobias Hoßfeld, and Kuan-Ta Chen. A survey of emerging concepts and challenges for qoe management of multimedia services. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(2s):1–29, 2018.

[17] Rob Van der Mei, Hans Van den Berg, Ivan Ganchev, Kurt Tutschku, Philipp Leitner, Pasi Lassila, Wojciech Burakowski, Fidel Liberal, Åke Arvidsson, Tobias Hoβfeld, et al. State of the art and research challenges in the area of autonomous control for a reliable internet of services. In *Autonomous Control for a Reliable Internet of Services*, pages 1–22. Springer, Cham, 2018.

[18] Aad Van Moorsel. Metrics for the internet age: Quality of experience and quality of business. In *Fifth International Workshop on Performability Modeling of Computer and Communication Systems*, volume 34, pages 26–31. Citeseer, 2001.

[19] Martín Varela, Patrick Zwickl, Peter Reichl, Min Xie, and Henning Schulzrinne. From service level agreements (sla) to experience level agreements (ela): The challenges of selling qoe to the user. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 1741–1746. IEEE, 2015.

[20] Qihui Wu, Guoru Ding, Yuhua Xu, Shuo Feng, Zhiyong Du, Jinlong Wang, and Keping Long. Cognitive internet of things: a new paradigm beyond connection. *IEEE Internet of Things Journal*, 1(2):129–143, 2014.

Paper B

Modeling Quality of IoT Experience in Autonomous Vehicles

Authors:
Dimitar Minovski, Chrsiter Åhlund, Karan Mitra

Reformatted version of paper accepted for publication in:
IEEE Internet of Things Journal, 2020

Modeling Quality of IoT Experience in Autonomous Vehicles

Dimitar Minovski, Christer Åhlund, and Karan Mitra

Abstract—**Today's research on Quality of Experience (QoE) mainly addresses multimedia services. With the introduction of the Internet of Things (IoT), there is a need for new ways of evaluating the QoE. Emerging IoT services such as autonomous vehicles (AV) are more complex and involve additional quality requirements, such as those related to machine-to-machine communication that enables self-driving. In fully autonomous cases, it is the intelligent machines operating the vehicles. Thus, it is not clear how intelligent machines will impact end-users QoE, but also how end-users can alter and affect a self-driving vehicle. This article argues for a paradigm shift in the QoE area to cover the relationship between humans and intelligent machines. We introduce the term Quality of IoT-experience (QoIoT) within the context of AV, where the quality evaluation, besides end-users', considers quantifying the perspectives of intelligent machines with objective metrics. Hence, we propose a novel architecture that considers Quality of Data (QoD), Quality of Network (QoN), and Quality of Context (QoC) to determine the overall QoIoT in the context of AVs. Finally, we present a case study to illustrate the use of QoIoT.**

Keywords—**IoT, QoE, QoS, AI, V2X**

I. Introduction

Internet of Things (IoT) represents a network of physical objects equipped with Information and Communication Technologies (ICT) connecting to the internet [1]. IoT enables advanced services for remote controlling, management, and maintenance of physical objects and processes [2]. Figure 1 shows a few examples of emerging IoT services in health-care, robotics, smart cities, and Intelligent Transportation Systems (ITS). It is expected IoT to connect billions of new devices to the internet [3] and therefore expands to services with a range of unique properties and quality requirements. Consider Autonomous Vehicles (AV) as an example, where the focus is on fast and accurate processing of sensory data to avoid collisions; while IoT-enabled remotely controlled vehicles impose stringent latency requirements on the machine-to-machine (M2M) communications. Such emerging IoT services are engaging closely to human life-style and privacy, where the delivered quality of the M2M type communications pose life-threatening risks, as well as risks of causing business losses.

From a stakeholder perspective (e.g. service providers, owners), service quality evaluation is an indispensable business goal for ensuring high end-users' satisfaction [4]. QoE was therefore introduced to measure the overall service acceptability, as perceived subjectively by the end-user [5]. However, the conventional QoE techniques mainly cover single end-user perception of multimedia services [5] and currently lack understanding of forming quality requirements for critical M2M type communications [6]. In particular, it is not clear the relationship between the end-users and the intelligent machines, such as an AV. A self-driving vehicle can simultaneously impact multiple end-users (e.g. passengers or pedestrians) and people in surrounding vehicles. Moreover, the self-driving can be triggered, adjusted, and overridden by a driver physically being in the vehicle or remotely from a distance. Herein, human-machine interaction (HCI) becomes increasingly complicated as multiple end-users can, in real-time, provide inputs and thus alter the decisions by the AV. Thus, policymakers are struggling in defining regulations and quality requirements for AVs, such as identifying responsible entities in cases of accidents involving AVs [7].

D. Minovski, C. Åhlund, and K. Mitra, are with the Department of Pervasive and Mobile Computing, Luleå University of Technology, Skellefteå, Sweden. D. Minovski is also with InfoVista Sweden AB. E-mails: *firstname.lastname@ltu.se, firstname.lastname@infovista.se*

Autonomous services, such as AVs, are typically orchestrated by Artificial Intelligence (AI), which enables intelligent data-processing, reasoning, and decision-making [8]. A single decision by an AV depends on various data-sources, gathered from multiple channels, creating complex interconnected chains. For instance, an AV combines the outputs from its internal predictive models, such as collision avoidance, line-keeping, detecting moving objects and signs in ensuring safe and efficient drive [9]. Each of the predictive models process data from the vehicle's hardware, but also can rely on data coming from the surrounding vehicles and road-objects [10]. Thus, a faulty outcome of a particular AI model, physical, or network object can lure the self-driving vehicle in making wrong decisions, potentially endangering human lives, directly impacting the QoE [11]. Therefore, in this article, we argue for an extension of the conventional QoE evaluation in understanding and modeling the behavior of intelligent machines. The present challenge is to identify the main components and logical processes of an autonomous service and measure their impact on the overall service performance. This is especially significant for businesses incorporating autonomous services, as they require real-time evaluation of the intelligent machines and their impact on business-oriented metrics, such as productivity, efficiency, and safety [2].

Contributions. In this paper, we define and model the concept of Quality of IoT-experience (QoIoT). Essentially, QoIoT represents an aggregation of metrics that describe the experience from two different perspectives: *(1) human*, by extracting the end-user's QoE perception [5]; and, *(2) machine*, by translating the overall machine performance to business-oriented metrics. As a consequence, the key contributions of this study are:

1) Exploring the complexities of autonomous processes and defining the term Quality of Machine Experience (QoME).

2) Extending the legacy QoE models in the IoT domain by defining the term Quality for IoT (QoIoT).
3) Proposal of a layered-QoIoT architecture, consisting of methodology for benchmarking IoT service performance.

The rest of the paper is organized in the following order: Section II presents the related work; Section III discusses the challenges in defining QoE for IoT and defines QoME; Section IV defines QoIoT; Section V presents the case-study; Section VI proposes the QoIoT architecture; and finally VII concludes the paper.

II. Related Work

This section reviews the current state of the art in the two main focus areas of this article: QoE for IoT; and Autonomous Vehicles (AV).

A. QoE for IoT

Traditional Quality of Service (QoS) parameters such as delay, jitter, packet loss, and bandwidth are used to objectively evaluate network quality. However, QoS narrowly captures the end user's perception of the service and may not reflect their overall service perception or experience [12]. To highlight the end user's point-of-view, the concept of QoE was defined as: *"The overall acceptability of an application or service, as perceived subjectively by the end-user"* [5]. Since then, the research on QoE has focused from mapping QoS metrics to users' QoE [13]; to understanding and defining factors influencing QoE [12]; and to building QoE management models [5]. Typically, the research on QoE is addressed from two distinct but often complementary perspectives, end-user applications and access networks [14, 15]. The network QoE management is mainly an extension of the legacy Quality of Service (QoS) techniques, measuring metrics to describe network performance and the way they impact the end-user perception [15]. Whereas application QoE considers social, cognitive, and contextual metrics to produce a score which ranks the end-users' experience [16]. The evolution of network technologies such as Software Defined Networks (SDN) and Network Function Virtualization (NFV) enable new QoE management mechanisms with dynamic orchestration and programmable resource allocation. While such technologies offer clear opportunities to improve the network performance, challenges arise in developing novel QoE monitoring methods due to the virtualization and lack of standards for orchestration [14].

The concepts of QoE and IoT are widely discussed in the research community [2, 5], however, mostly in isolation. The combination of QoE for IoT is still in its infancy due to a lack of research that clearly distinguishes the emerging IoT services (see Fig 1) from multimedia in terms of their impact on QoE. Wu et at. [1] presents one of the earliest works discussing QoE for IoT. They propose the concept of Cognitive IoT as a new network paradigm where physical and virtual objects are interconnected and perform as agents. In addition, Wu et al. [1] identified layers of data in an IoT service to be considered for QoE evaluation. Namely, they distinguish a data layer from information layer, where the former one aims to evaluate the quality of the sensed data, while the latter one deals with a contextual meaning of the data. The study [1] also presents research challenges facing QoE assessment methods in the domain of IoT. Among their findings, they identify: (1) A clear challenge of obtaining high quality data with mixed characteristics and massive amount of sources; (2) Current QoE state of the art is limited on a single user QoE, failing to cover system-level QoE suitable for large scale IoT services with massive user base; and (3) A clear need for developing effective semantic and knowledge discovery technologies that are suitable for supporting large scale IoT services.

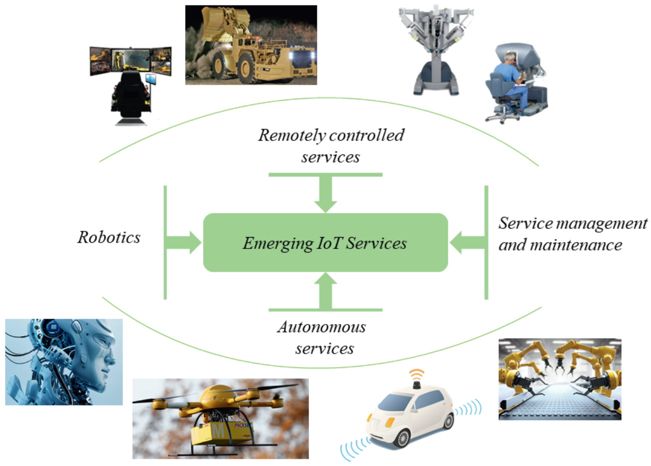


Fig. 1: Emerging IoT services.

Floris and Atzori [17] takes a step further in developing a concept of a layered-QoE framework to evaluate the quality of a multimedia IoT service. The validation of the model is performed with gathered data from an IoT vehicle application, containing sensors and camera data. The data-sources are streamed and displayed to end-users, which then subjectively assign a QoE rating score. Within their model, Floris and Atzori [17] propose the concept of Quality of Data (QoD) for benchmarking the accuracy of the produced IoT data. Although the authors do not provide methodological steps on what it takes to measure QoD, they use an abstracted version of it in combination with QoS to map to subjective QoE scores. However, the studies [1, 17] provide no indications on how the models could cover other IoT services except multimedia, shown in Figure 1. QoE models for multimedia services typically evaluate the perception of a single user, while AVs, for example, expand into multi-user interactions with drivers, passengers, pedestrians, and other drivers on the road. Our study builds upon the works of [1, 17] by significantly expanding the discussion in the direction of: (1) end-users' interaction with AI and autonomous services; and (2) evaluating the quality and performance of an autonomous service.

B. Autonomous vehicles

The concept of the Internet of Vehicles (IoV) recently emerged as a result of the advances in sensor technology, artificial intelligence (AI), and wireless networks, as well as the customer needs for AVs [18]. As a sub-group under the IoT umbrella, various AVs are envisioned to posses abilities for dynamic communications, storage, and intelligence to anticipate customers' intentions. Few architectural IoV proposals already emerged, among which is Cisco's with four layers [19]: (1) End-point, containing the vehicle's hardware and software; (2) Infrastructure, defining all technologies that allow connections; (3) Operation, to monitor the policy enforcement and the flow-based management; and (4) Virtual layer, with all the different types of cloud subscriptions. Other IoV architectures are surveyed by Contreras et al. [20] from network protocols and security perspective, while Ang et al. [18] reviewed IoV architectures for a smart city environment and applications. What unifies those architectures is the very little attention given to the quality evaluation of the AV as a complete end-to-end service and the lack of discussion on data-driven methodology to monitor the QoE. A typical AV case integrates three technological visions [20, 21]: (1) Human-Computer Interaction (HCI); (2) Intelligent processing; and (3) Networking. Their relationships and interactions, illustrated in Figure 2, are incredibly complex [21] and before understanding the quality concerns within, we first need to address the current state of the art in each of the three areas, summarized in the following list:

1) Human-Computer Interaction (HCI): The research on HCI in ITS has focused on the way end-users interact with the automotive interfaces, primarily the Advanced Driving-assistance Systems (ADAS) [22]. However, the scope of HCI in the context of AVs enlarge as the AI models first need to understand and learn from human behavior in driving a vehicle. Xing et al. [23] conducted real-life experiments for driver activity recognition; while Li et al. proposed a rather cost-effective solution [24] that utilizes a driving-simulator to train and test the AI models. Next, one significant challenge is how the human will understand the AI, in particular, of a self-driving vehicle. The HCI models are, in this case, multi-dimensional, as the AI system interacts with the driver, passengers, pedestrians, and other drivers on the road [25]. For instance, popular self-driving commercial vehicles require full attention of the driver [9]. In this regard, a large-scale study by Fridman et al. [9] tracks the driver's behavior and interactions during self-driving to interpret the circumstances when the driver overrides a decision by the AI. However, the interactions between an AV with other passengers, pedestrians, or drivers on the road are still not clear; thus, novel HCI models are needed to understand the role of the human in affecting and altering the autonomy.

2) Intelligent processing: AVs aim for driving autonomy where the decision making is based on the vehicle's hardware, ranging from Lidar, GPS, video cameras, to CAN Bus sensors that monitor the vehicle status [26]. That is, AVs aim for full self-driving, with or without a human on the driver's seat. AI and Machine Learning (ML) techniques have been utilized in providing the needed intelligence when processing the vehicle's data. An approach of Lidar, Radar, and Camera fusion using reinforcement learning (RL) is proposed by Rangesh et al. [26] to track moving objects in real-world driving conditions; while Li et al. [27] used Convolutional

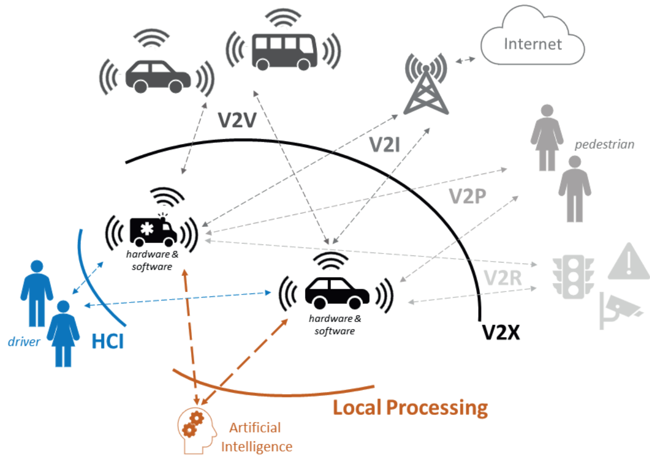


Fig. 2: Technological visions of autonomous vehicles.

Neural Networks (CNN) for dynamic vehicle detection on the road. The predictive outcomes from such models are combined with outputs from traffic sign detection and line-keeping methods for predicting and avoiding collision on the road [28, 29]. However, having individual models for different tasks raises quality concerns in the research community [21]. In particular, how multiple interconnected predictive models will co-exist together, tolerate each other's failures, and impact the overall vehicle performance to ensure safe and optimal drive [21]. In addition, the use-cases for AVs range from replacing conventional cars to revolutionizing industries with autonomous logistics, UAVs, mining and construction vehicles with self-driving and autonomous task execution. In the case of mining vehicles, for instance, the autonomous software needs to consider and understand the impact of its real-time decisions on business metrics, such as production levels and reliability [6].

3) Networking: Communication among vehicles is considered to be an important factor ensuring safety by gathering relevant environmental information from the surroundings to predict events before they occur [10]. In that direction, the 3GPP has been actively extending the requirements for future cellular wireless networks to support AVs connectivity, known as Vehicle-to-everything (V2X). An LTE based V2X standard has already been issued with 3GPP Release 15 [30], while the new 5G/NR eV2X standard is envisioned for the end of 2019 with Release 16 [31]. Besides cellular connectivity, ETSI and IEEE have been active with the development of 802.11p protocol to support V2X [10]. A survey by Cheng et al. [10] reviews the options of wireless technologies for IoV. Figure 2 depicts the interaction models for V2X, while the following list summarizes their main use-cases, based on the 3GPP and 802.11p requirements [30, 31, 10].

- Vehicles-to-Infrastructure (V2I): includes infotainment (information and entertainment) services, such as media streaming, VoIP, and web-browsing, but also road-map downloads and software updates [25]. V2I also enables remote-supervision of AV, in industrial cases, for example, where the vehicle can also follow remotely given orders.

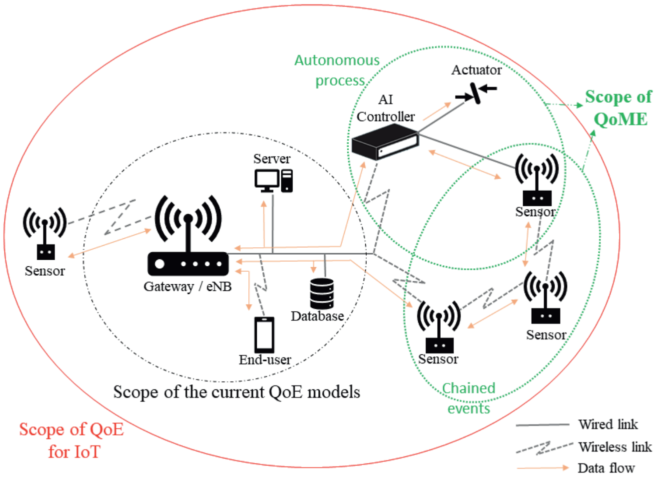


Fig. 3: The scope of QoE for IoT.

- Vehicle-to-Vehicle (V2V): includes real-time cooperative awareness message (CAM) exchange among surrounding vehicles on safety and road planning [25]. For instance, warnings on a slow/fast vehicle, collision risk, and speed limits. It also includes decentralized environmental notification messages (DENMs), with requirements of maximum 100ms latency, such as vehicle accident, traffic conditions, hazardous location, visibility, precipitation, and road adhesion [25].

Besides the two main interaction models, V2I and V2V, 3GPP defines requirements for Vehicle-to-Pedestrian (V2P), ensuring pedestrian safety, and Vehicle-to-Roadside (V2R) for communicating with static road-objects and traffic signs [31]. However, having an abundance of external communications at the same time infers security threats [20], but also makes the AVs vulnerable from a common set of wireless network issues, further degrading the QoE. For instance, the dynamic environment generates issues with the wireless channel estimates and doppler effect, resulting in interference and poor signal detection. In the future, most of the vehicles will have internet connectivity [25], and today's LTE networks are already facing scalability issues [32]. Congested cells create additional latencies and packets drop, which can be devastating for AVs. For instance, a self-driving decision to overtake a vehicle can depend on messages received via V2X interfaces, and inaccurate or delayed packets can lure the AI in making a wrong decision [25].

III. Challenges in Defining QoE for IoT

This section categorises the main challenges in evaluating QoE in an IoT service, based on the state of the art research in QoE and AV. It discusses the complexities of autonomous services and more importantly, why IoT requires a revised version of the conventional user-centric QoE.

A. System of Systems (SoS)

A fundamental characteristic of an IoT service, especially in the industrial domain, is the coalescence with functionalities provided by internal and external services, with input and output data coming from both humans and machines [2]. This has led to complex service chains, known as System-of-Systems (SoS) [2], with possibly even hundreds of services offered by different third parties, each with their own business incentives. Consider Figure 3 as an example of an emerging IoT service, where a single decision by an autonomous process may be taken as a result of multiple data sources that are dependent on one another. Now consider a mining company, with a fleet of AVs digging and extracting the ore. The AVs periodically receive CAMs and DENMs updates from multiple sensors that detect smoke, humidity, air pollution, and temperature. In case of an emergency, the risk assessment of the safety alarm service triggers an alarm, which is followed by a series of chained events, including locating the miners and AVs and computing their most optimal rescue route. Then, the service must promptly notify the miners and AVs for their own rescue routes and contact additional external services to deal with the emergency, such as the fire department in case of a fire. Potential sources of failures in complex SoS are multi-dimensional, arising from: (1) IoT hardware and software, with respect to the quality of the produced data [33]; (2) Network, related to the latencies and packet losses due to, for instance, dynamic wireless environment, firewalls, encryption, congestion, interference, virtualization [25]; (3) Virtual computing, with respect to data processing, caching, and data storage [34]. In current practice, the quality evaluation of complex SoS is controlled on an ad-hoc basis, while the consequences of failures are not well understood [6]. Therefore, there is an urgent need for QoE measurement and monitoring regarding complex SoS offering possibilities for root-cause analysis [6].

B. Quality of Machine Experience (QoME)

Figure 2 illustrates the main components and interaction models of the future AVs, with AI being the key enabler. Going into the essence of AI, it is defined as *"the science of making machines capable of performing tasks that would require intelligence if done by humans"* [8], while others refer to AI as *"similar or the same kind as human intelligence exhibited machines"* [21]. The research on AI goes as far as defining machine emotions [35] and human-robot interaction [36]. Thus, one may envision that the AI-enabled controller (Figure 3) may, just like humans, perceive *experience*, which arguably is different compared to the humans (e.g. QoE). AI is trained with data-driven approaches [8], and recent research in Explainable AI (XAI) suggests that any decision by the AI is traceable and explainable with objective data and mathematical formulas [37]. Therefore, the perceived quality and experience by the AI during run-time can be evaluated with data-driven approaches, and defined as Quality of Machine Experience (QoME) as: *"Objective metric that measures the quality and performance of intelligent machines and their decisions."*

A precursor to full autonomy is knowledge creation and reasoning capabilities on the collected information from the environment [6]. Herein, the presence of the human entity is important in providing labels to train the models. More advanced AI techniques that enable full autonomy integrate self-learning techniques [38], that consist of advanced models of the world, pattern recognition of novel events, and self-reliant reaction in unexpected events. Self-optimization is another logical process that enables full autonomy. It means that the AI model uses benchmarking metrics while doing a particular task to fine-tune its operations according to the benchmarks [8]; avoiding the classical *trial and error* AI concept [8]. In this direction, we introduce QoME as an objective knowledge used by the AI for self-optimization. Consider an AV overtaking a vehicle on the road. Such a decision can depend on the settings defined by the end-user, as well as on the input from a variety of internal and external sources; for instance, camera, lidar, and CAMs/DENMs from surrounding vehicles via V2X communication. The data is then processed in real-time by various predictive models that produce outputs which are inputs to the other models. For instance, a predictive model for overtaking a vehicle reasons on inputs from models such as detecting moving objects and traffic signs, line-keeping, collision avoidance, and speed control. Thus, the M2M interaction among those predictive models is itself a machine experience that can be objectively benchmarked.

The scope of QoME is supervision and benchmarking of such predictive models in their contribution to the overall system performance, productivity, and safety. To do so, supervising AI models can be used to detect deviations from normal workflow (e.g. condition monitoring [8]), find root causes for such anomalies (e.g. diagnostics [39]), calculate optimal parameters, and predict the future system behavior (e.g. predictive maintenance, energy optimization [8]). In that direction, the critical challenge remains in identifying the main data-sources in a service (e.g. IoT hardware, network, logical AI processes) and model them to produce and monitor QoME metrics.

C. Summarizing the main challenges regarding QoE for IoT

The following list exposes the limitations of the current state of the art methods for monitoring and evaluating QoE for IoT, providing questions to which the current literature does not have well-defined answers:

1) How can an end-user alter or affect the performance of a complex SoS, that also includes autonomous processes?
2) How to evaluate the quality of autonomous processes and troubleshoot their impact on the overall service performance?
3) With what type of objective knowledge for quality evaluation and control to feed the self-optimizing methods in an autonomous service?
4) How to perform a root-cause analysis in complex SoS, containing intelligent machines and multi-user interactions?

IV. Quality of Internet of Things-experience (QoIoT)

In this section, we present the extension of the user-centric QoE paradigm to cover emerging IoT services by defining Quality of IoT-experience (QoIoT), addressing the challenges

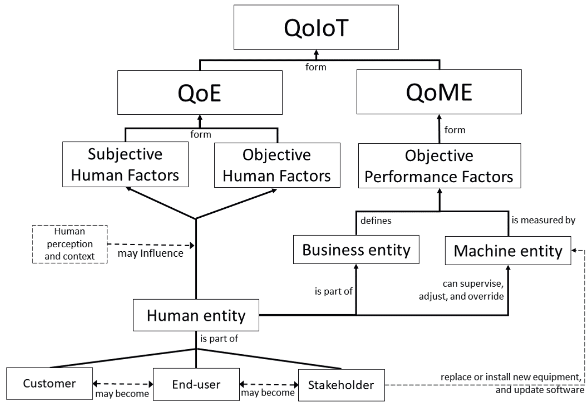


Fig. 4: Defining QoIoT-experience [40].

from Section III. Essentially QoIoT is a combination of metrics that are measured from two different perspectives: (1) QoE, by extracting metrics from the legacy user-centric management models [12, 14, 15]; and, (2) QoME, by understanding machine's behavior and translating the overall machine performance to business-oriented metrics. A high-level overview of QoIoT is illustrated in Figure 4, and can be represented as:

$$QoIoT \in \{QoE, QoME\} \tag{1}$$

Correspondingly, the newly formed definition of QoIoT states as follows: *"Quality of IoT-experience (QoIoT) is a metric that aggregates the delivered quality of an IoT service from the perspective of humans and machines. The end-users' experience is evaluated with subjective and objective human factors, while Quality of Machine Experience (QoME) translates the overall machine performance to business-oriented metrics through objective performance indicators."*

Figure 3 shows the high-level differences between QoE and QoME. Within end-to-end QoIoT evaluation, we envision QoE and QoME to leverage each other in improving the overall service performance. For instance, an end-user in a typical multimedia service watches a video stream via the internet, downloading frames from a remote video server. Therein, the main factors affecting the QoE are the network performance (e.g. QoS), subjective (e.g. mood, enjoyment, emotions) and objective human factors (e.g. Human visual system (HVS)). Such metrics are still important for autonomous services, as the end-user may remotely-control a machine via video live-stream. However, in cases of full autonomy, a driver can supervise the vehicle during self-driving, where the QoE is now affected mainly by the performance of vehicle's hardware and AI-powered predictive models; thus creating a gap in QoE evaluation as it is difficult to measure the reasons behind poor performance. We envision QoME to bridge this gap by benchmarking the autonomous processes to improve the end-users' understanding and perception of them. From a human perspective, QoME essentially measures the impact of machines (e.g. full autonomy) on end-user's QoE. Showing QoME metrics, such as the status of the safety, to the end-user is especially important during self-driving, as it can improve the driver's situation awareness to promptly adjust

Components in an IoT service	Key Performance Indicators	Classification
Physical	*Hardware type (e.g. sensors, actuators, cameras, lidar), CPU, OS, raw data, labels, memory, display resolution, data-type format, sampling frequency, sensing characteristics, energy consumption, cryptography [2]*	QoS
Network	*Jitter, packet-loss, bandwidth, delay, buffer size, Received-Signal-Strength Indicator (RSSI), Reference Signal Received Quality (RSRQ), Signal-to-Noise Ratio (SNR), etc. [10]*	
Application and service specific	*Application type (e.g. video, voice, IoT, web), Application metrics (e.g. bit-error rate, encoding, bitrate, content genre) [5]; Subjective human scores (e.g. MOS, mood, enjoyment, emotions, easy-of-use, end-user's characteristics and background), Objective human scores (e.g. HVS) [12]*	QoE
Logical system components	*Benchmarking individual data-sources (e.g. anomaly and novelty detection, context-aware pattern recognition [39]), benchmarking predictive models (e.g. condition monitoring, diagnostics, predictive maintenance [2]), benchmarking the impact of autonomous processes on overall service performance (e.g. productivity, safety, reliability)*	QoME QoIoT

Fig. 5: Differences and synergies among QoS, QoE, and QoIoT.

the vehicle's settings or completely takeover and stop the self-driving.

Future AVs are intended to operate without a human being in the vehicle [9], where machines alone can have a devastating impact on their surroundings and also affect the business. In the following, we further explore the way QoME, and thus QoIoT, can leverage autonomous services through analysis of business-machine and human-business relationships, part of Figure 4.

A. Human-Business Relationship

QoIoT, by definition, extends the user-centric QoE approaches and therefore extracts metrics from the Human Entity (Figure 4), composed of end-users, customers, and stakeholders [12]. The business entity, driven by the customer needs defines Service Level Agreements (SLAs) for ensuring high quality of their services; based on which the business operations and goals are optimized. Varela et al. [4] proposed an enhanced version of SLAs, which includes evaluation of QoS and QoE metrics to improve the business-to-customer relationship. Figure 5 summarizes the synergies among QoIoT with the current state of the art, classified based on the target groups for each quality evaluation concept. The first criterion in Figure 5 is QoS, which strictly describes the network quality with objectively measured metrics [12]. Next, the QoE management models [14, 5] combine QoS metrics with subjective and objective human factors to evaluate the end-user's experience of using a particular application (e.g. video, voice, web).

In the presence of intelligent machines, as discussed earlier, QoME objectively evaluates their performance. In that direction, measuring conventional QoS/QoE metrics can leverage the evaluation of QoME in: (1) Learning/training phase – an AI system typically learns from human behavior and thus, measuring subjective and objective human metrics is beneficial for understanding the human impact on the QoME metrics; (2) Run-time phase – measuring subjective and objective QoE while, for instance, a driver overrides a self-driving AV can leverage the AI self-learning techniques in labeling the edge cases where the AI is not performing well.

B. Business-machine Relationship

Endsley [11] defines the term Situation Awareness (SA) in dynamic systems; it means to observe the current state of the system, detect patterns and create knowledge that improves service control and performance. Therefore, in complex autonomous services it is essential to create QoIoT knowledge understandable for both entities, human and machine (Figure 4). On the one hand, the Business entity needs such knowledge to: (1) Understand the impact of the machines on the business and user-centric QoE, enabling the stakeholder to dynamically adjust and prioritize certain QoIoT metrics; (2) Replace or install new equipment, and upgrade software; (3) Semi-control the machines; and (4) Adjust the settings of autonomous machines. On the other hand, measuring QoIoT is beneficial for the Machine entity regarding [6]: (1) Understanding the impact of individual data-sources and AI logical components on the overall service performance; (2) Using the QoIoT metrics as benchmarks for optimizing the autonomous operations, for instance by utilizing RL.

QoIoT considers metrics depending on the business context and the case study. Figure 5 gives an example of conceptual factors applicable to the AV. The way QoIoT metrics are defined will depend on the business goals and service requirements; thus, the stakeholder is the prime actor in their formation. In addition, the QoIoT metrics will depend on the level of machine intelligence, network usage, presence of end-users, and their interactions within the service. The next Section V provides an example of forming QoIoT metrics around a case-study involving AVs. Then, in Section VI, we identify the main data-sources in an AV service, based on which we build a methodology for evaluating QoIoT, for instance, measuring the progress of a particular business objective, such as productivity, safety, and reliability. The book by Samsonowa [41] describes several examples of other QoIoT metrics in various ICT fields. They can be rather general, such as Overall Equipment Efficiency (OEE) [41] or industry-specific, such as revenue/hour.

V. Industrial IoT Case Study: autonomous vehicles in mining

Autonomous mining vehicles with the possibility of remote control are considered as a case study to apply the proposed QoIoT concept and define metrics for monitoring. The motivation is to automate the operations in mine, such as blasting and extracting ore, and remove the driver from the hazardous underground environment. The aim is to utilize the production hours of the mine, especially during lunch breaks, shift changes, and blast hours - as the mine workers typically wait few hours after a blast to access the mine due to polluted air [42].

The mining IoT setup is illustrated in Figure 6, consisting of a variety of sensors, actuators, and cameras, streaming V2I data in uplink over the cellular network to a remote-control station. An expert-driver can supervise and control

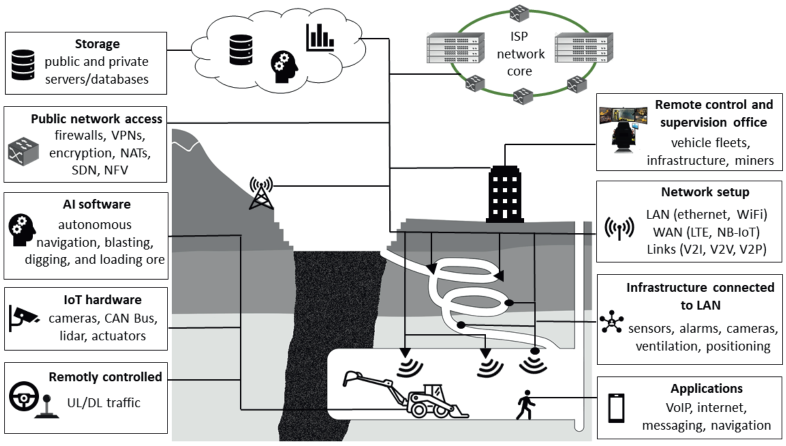


Fig. 6: IoT setup in mining.

the machines from a distance, where data is being sent in downlink as well, from joysticks, buttons, and pedals. The variety of bi-directional V2I traffic sources impose complex quality requirements for the network, as the traffic consists of critical and non-critical information, substantially and lightly weighted data. In addition, the communication in the mine (e.g. during a remote-control, VoIP, internet, and data storage in a database) must go via the nearest network core (Figure 6). This is important because all the packets travel back and forth from the mine to the core network, creating a large overhead [10]. Then, the added security measures shown in Figure 6, such as firewalls, VPNs, encapsulation, and encryption additionally complicates the network quality evaluation [14].

Each mining vehicle can also execute tasks autonomously relying on an AI software, as shown in Figure 6. For instance, the vehicles can independently drive in-and-out of the mine, evacuate in case of emergency, automatically place and perform the blasting, and assist in loading the ore. Being autonomous means that specific hardware and software is deployed on the vehicles capable of decision making and controlling the vehicle. As the operations are orchestrated purely by centralized software, the run-time of the system must integrate complex processes such as self-learning, self-optimization, and self-healing [2].

A. Defining objective QoME performance indicators around a case-study

The mining industry has predefined metrics which are continuously measured, based on which the business operations are optimized; Andersson [42] identified industrial revenue, safety, and productivity as Key Performance Indicators (KPIs) within mining. While other businesses involving AVs generate revenue from reliable and efficient service, in the mining case-study it is the productivity as the main factor determining the revenue. The productivity as a metric is a function of the productive capacity meeting its intended end-results, such as planned production levels, availability, extracted tons/h, operational and down-time hours [42]. Measuring the productivity, in this case, requires behavioral monitoring of the expert-driver and individual data-sources from the IoT setup in Figure 6. Such role separation enables the possibility for root-cause analysis when the productivity levels are not at the desired level. A real-time productivity metric can enhance the expert-driver's SA during remote-control or supervision of the AV, regarding enabling/overriding a self-drive or adjusting the settings of the AI. Having the expert-driver intervene during self-driving or adjusting the settings is beneficial for the AI system itself, as the human behavior and the productivity metric can be used to label edge cases. Moreover, it enables the AI to adapt to changes by the stakeholder. At times, it may be necessary to decrease the production levels and increase safety, and the AI system must self-optimize its operations according to the requirements.

The productivity in mining case can be defined with the following QoME metrics:

$$\begin{aligned} Productivity \leftarrow Production\ level\ \vee\ Reliaiblity \\ \vee\ Efficiency \end{aligned} \tag{2}$$

The following list explores the ways to monitor the objective QoME KPIs, relevant to the mining case study:

1) *Production level:* is defined as loading the maximum possible scoop of ore in the least amount of time, rendered as a tons/hour [42]. Herein, the essential data-sources need to be identified, which leads to the desired tons/hour levels. For instance, weight, vibrations, trajectory, and pressure on the filling bucket can measure scooped tons/hour. A real-time ML analysis on such metrics, combined with CAN Bus messages, can compute the most suitable vehicle's acceleration while excavating the ore [43]. In a similar way, the most optimal technique to load as much ore as possible per scoop can be computed, depending on the rock structure, directly increasing the tons/hour metric. In cases when an expert-driver is remotely-controlling the vehicle, a real-time analysis on the network QoE, as well as video streaming QoE [14], can boost the overall productivity. In our previous work, we have studied the degradation of video streaming quality during real-life drive tests in a mine [44]. A video Mean Opinion Score (MOS) was predicted using passive in-smartphone measurements that describe the wireless LTE link, such as Signal Strength Indicator

(RSSI), Reference Signal Received Power (RSRP), and Signal-to-Interference-Noise-Ratio (SINR). A forecast of poor video quality can suggest to shut-down one of the live-streaming cameras on the vehicle to enable seamless operations; but also to completely switch to full self-driving when the network cannot support heavy live-streaming.

2) *Reliability:* refers to the direct costs of repairing a broken vehicle. However, it can also be associated with replacing or updating parts of the remote-control setup, hardware or software, anticipated from the subjective human factors, such as easy-of-use and technology acceptance. Ghazizadeh et al. [45] measured the driver's perception of the technology, where poor acceptance has a direct impact on the productivity. The maintenance cost of the vehicles, autonomous software, and the network infrastructure include production losses during the downtime period, also directly impacting the overall productivity. The reporting of issues regarding parts of the IoT setup (Figure 6) can be both subjective and objective. Besides a human directly experiencing issues with the system, objective models can be used for directly predicting issues before they occur. Herein, thorough data analysis is necessary for anomaly detection on the data-sources, involving self-supervised ML, such as data clustering and Principal Component Analysis (PCA) [6]. An outcome can predict when is the right time to replace hardware or upgrade software, before a failure occurs. Dadhich et al. [43] surveys predictive AI models for common reliability threats in mining vehicles, such as wheel slip and collision avoidance. For instance, Nam et al. [46] use CAN Bus messages to develop a traction control system in avoiding wheel slips. Similar data analysis can also measure network reliability in anticipating failures during a remote-control. In one of our previous studies [47], we have used LTE link characteristics, such as RSSI, RSRP, and SINR, to predict one-way latency. Therein, a root-cause analysis on a predicted high latency anticipates network issues, such as high load on the cell, handover, and high interference. Having a passive, non-intrusive, estimation of the latency can also anticipate upcoming blind network coverage spots, where the expert-driver would lose a connection with the vehicle.

3) *Efficiency:* is associated with the success rate of correct operation, mean execution time, rate of operating errors, and complexity of the currently executing task. Fuel/battery efficiency is the most obvious metric in vehicles. For instance, data-sources such as CAM Bus (e.g. gear ratio, engine inertia), battery characteristics, windward area, and air resistance are modeled in [48, 49] using RL for power management. Andersson [42] conducted a large study that groups trajectory mappings, positioning data, and V2V messages to compute the most efficient driving path in mines. The same study also identified data-sources for monitoring other efficiency metrics for mining vehicles, such as bucket payload (tons), speed (km/h), transport performance (ton*km/h), and machine-utilization (%).

VI. Model, measure, and predict QoIoT

The focus of this section is to describe an architecture for data-driven evaluating QoIoT by following three pillars – Model, Measure, and Predict, as suggested by Mitra et al. [12]. The idea of having the three entities is eliciting clear separation of the roles within the quality evaluation process. That is, the proposed architecture attempts in addressing the challenges from Section III by: (1) Identifying and model the data-sources in an IoT service; (2) Identifying the way end-users' interact with the service; (3) Identifying QoIoT metrics for monitoring the logical processes of intelligent machines; and (4) Forecasting the service performance and providing a root-cause analysis of poor KPIs. Leveraging each other, the three entities utilize specific goals in quantifying the human-business and business-machine relationships. The idea is to identify KPIs for measuring QoIoT that would be understandable for both humans and machines. Thus, the aim of QoIoT is to serve as a transparent bridge between: (1) The physical world, including the IoT objects and resources; (2) Virtual world, such as cloud computing, data transmission, and intelligent processing; and (3) Social world, translated as end-users' experience and business needs.

The architecture of QoIoT is illustrated in Figure 7. In the following, each of the three entities is described in detail, utilizing autonomous vehicles as a case study.

A. Model QoIoT

The performance of complex SoS depends on the interactions between the control system (e.g. centralized software, predictive models), physical (e.g. IoT hardware), and virtual components (e.g. data transmission, cloud computing) [20]. Thus, the new ISO 26262 standard on functional safety in road vehicles recommends such interactions to be considered for evaluation during the run-time of the service [50]. Figure 7 shows the prerequisites of input parameters for the model. Live access to raw data, bitstream, and network are dynamic information that the model reasons on. It includes the data produced by the IoT objects and logical components, as well as their communication. A general information and service requirements are static initial knowledge describing the case-study. For instance, it consists of specifications regarding the utilized hardware and software, network connectivity and protocols, and virtual components for data storage and processing. The model also needs to understand the service quality requirement of what is expected from the utilized control system, physical, and virtual components. However, addressing the issues of extracting the initial knowledge from the static inputs is out of scope for this study, and the assumption is that such information is already available in a standardized format ready for further analysis. In addition, QoIoT does not consider other safety recommendations from the ISO26252 [50], such as formal methods and hardware/software code checking, which we consider to be already conducted before service deployment.

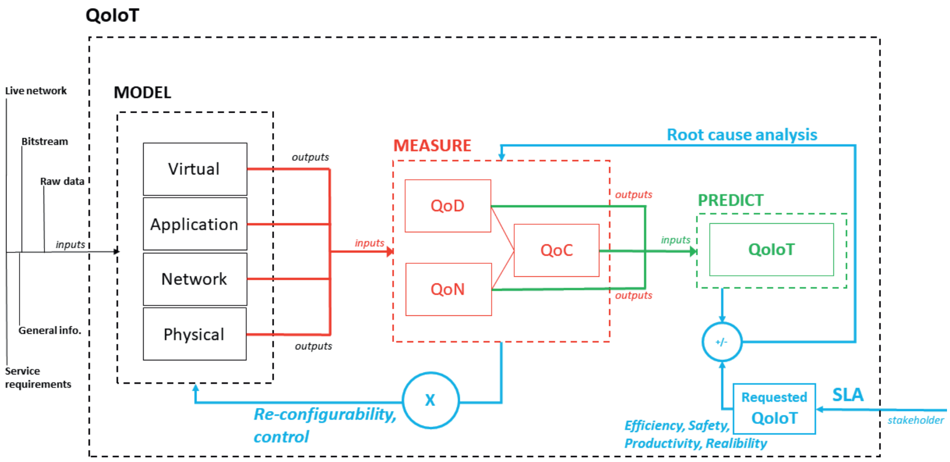


Fig. 7: QoIoT model, measure, and predict.

The QoIoT model consists of four layers, each with a predefined scope and set of methods to aggregate and process the input parameters. Generally, the model is responsible for the perception-action cycle, as the most primitive cognitive method, targeting physical, network, application, and virtual layers. The scope of the model is defined as: (1) Retrieve and understand the specification of the service requirements; (2) Detect the components of the service; (3) Monitor the state of the components by gathering as much real-time data as possible; (4) Control and re-configure the physical, network and virtual components. In the following, the classification of these functions is described in detail, covering each layer separately.

*1) **Physical Layer**:* As the name suggests, this layer is responsible for discovering the physical objects and monitoring their interactions within the system and environment. For instance, it consists of data acquisition from the vast number of physical objects in an AV, such as lidar, cameras, radar, CAN Bus (e.g. engine control unit, transmission, airbags, antilock braking, audio and battery systems). The layer requires an initial knowledge of the type and number of physical objects in the system, as well as their specification and service requirements. The initial knowledge enables the labeling of the created raw data, which is beneficial for the antecedent layers, especially when looking for anomalies in the produced data. Thus, the physical layer prepares a set of output metrics for further analysis, describing each physical object. For instance, type of objects, produced (labeled) data, raw data, sampling frequency, energy consumption, cryptography, multimedia/sensing features. One example may consist of the following layout:

- *Input options* = { *number of objects, hardware and software specification of the objects, etc.* }
- *Output metrics* = { *number of detected devices, hardware and software specification of the devices, type of produced data (boolean, binary, charachter), labeled data (vehicle speed, fuel level, lidar, camera image, etc), raw data, sampling frequency, energy consumption, cryptography, etc.* }

*2) **Network Layer**:* Communication networks, especially in AVs case, exhibit performance and reliability degradation caused by the dynamic wireless environment (e.g. coverage, doppler, interference) [10]. As a result, several deficiencies may arise, such as delays in data transmission, jitter, packet loss, and bandwidth congestions [12]. The scope of the Network layer is to monitor the connectivity and data transmission in the service and prepare metrics that will describe the state of the network.

An AV integrates inter and intra communication, where the former typically connects the physical objects with wires, while the latter one refers to the previously mention wireless V2X communication. Intra communication is important since it enables cooperation among the AVs to understand the greater context and prepare the vehicle in advance to avoid potential danger [10]. Kim et al. [51] demonstrate the importance of V2X communications, where poor or no connectivity leads to road accidents in AV. Multiple studies are reviewing the potential candidates for enabling V2X, namely 802.11p and LTE/NR-V2X [10, 25, 20]. Chen et al. [10] provided a detailed comparison based on use-cases for those protocols, with LTE-V2X emerging as more efficient in terms of packet delivery ratio, scalability, and mobility support. The requirements for the communication protocols in V2X range from high throughput in V2I case (infotainment and remote-controlling), to high reliability, low latency and low throughput for V2V case (CAMs and DENMs) [25]. The dynamic wireless environment requires modifications of the V2X protocols compared to the conventional LTE [31]. For instance, it is necessary to mitigate Inter-Cell Interference Coordination (ICIC), doppler effect, and have more efficient maintenance of routing tables to increase the cell capacity. In addition, the emerging 5G/NR standard promises enhancements to enable V2X, with massive Multiple-Input Multiple-Output (MIMO), direct device-to-device (D2D) communication, and advanced OFDMA transmission schemes [31]. Thus, understanding and measuring metrics that describe the new possibilities for V2X communications are the essence of the network layer.

Real-time end-to-end evaluation of the network is typically done by inspecting the packet-header and encoded bitstream (e.g. sequence number, RTP marker bit usage), as suggested by

ITU-T Y.1540 recommendation [52]. These measurements are performed during a well-defined time periods, referred to as observation window, which size is a trade-off between the robustness of measurement and responsiveness to changes [53]. Different types of probing messages are recommended by ITU-T Y.1540 [52], such as sending a ping or unbiased message sample to calculate delay and jitter. More complex network evaluation systems utilize bulk transfer tools to calculate the available throughput. For instance, SCReAM [54] and Coupled CC [55] measure congestion in the network by streaming the maximum amount of traffic during an observation window. Having observation windows is, however, a limiting option for AVs case due to the dynamic environment, requiring constant measurements. More cost-effective techniques involve monitoring passive, non-intrusive, metrics that can be used with ML models for estimating network performance. We will explore these options in the Quality of Network (QoN) layer, discussed in Section VI-B2. Typical non-intrusive metrics describe the wireless access link, such as RSSI, RSRP, RSRQ, and SINR.

To be able to measure all of the above-discussed metrics, the Network layer requires initial knowledge from the planning and design of networks, as well as service requirements regarding network quality expectations. Standard examples of such knowledge are described in ITU-T G.107 [56] and G.1070 [57], for speech and videophone, respectively. An example of an initial knowledge can describe the following characteristics: (1) Path direction and type of messages sent in uplink and downlink; (2) Bottleneck-link capacity, defining minimum capacity of the end-to-end path; (3) Bottleneck queue type, and bottleneck size in terms of queuing time when the queue is full; (4) Maximum end-to-end jitter; (5) One-way propagation delay; (6) Convergence time; (7) Path loss ratio, characterizing the non-congested, additive, losses to be generated on the end-to-end path; (8) Type of modulation; (9) Utilized frequencies and bandwidth; (10) Maximum and minimum data-rates; and (11) Utilized communication protocol with its characteristics and use-case.

As a summary, the proposed input and output metrics from the Networking layer are the following:

- *Input options* = { *Access to live network, network planning and design; observation window; type of measurements; packet-header and bitstream information; throughput, delay, jitter and packet loss ratio requirements; utilized frequencies and bandwidth; etc.* }
- *Output metrics* = { *Measured jitter, delay, and packet loss percentage; network bottleneck bandwidth; queue length; queuing delay; feedback message overhead; handshake time; measured interference, RSSI, RSRP, RSRQ, and SINR values; etc.* }

*3) **Application Layer**:* The scope of the Application layer is the end-user placement in the service and its experience of interacting with the hardware and software in the AV. Therefore, this layer requires initial knowledge of the end-user's role and interactions in the service. For instance, a driver in the vehicle can interact with it through voice, touch screens, buttons, pedals, and wheel. Whilst an end-user can remotely control and supervise an AV via live video streaming. Assessing live-video streaming has been surveyed in [14, 5]. ITU has addressed video QoE in multiple standards, where the most recent P.1203 recommendation [58] targets real-time video. P.1203 relies on information retrieved from the packet-header and bitstream, such as encoding, resolution, bit-rate, frame-rate, and frame number to calculate a QoE MOS. Even though P.1203 uses objective metrics to evaluate the QoE, the produced score is based on end-users subjective ratings.

Understanding of the subjective and objective human factors is important as they are also used in evaluating HCI [23]. Möller et al. [5] discussed a wide range of human QoE factors, including: (1) Subjective human factors - mood, enjoyment, emotions; (2) Objective human factors - Human Visual System (HVS); (3) Usability subjective parameters - easy-of-use and technology acceptance; (4) Contextual subjective parameters - end-users' characteristics, social and cultural background. Within the vehicle industry, the evaluation of human QoE factors (Figure 5) so far has been limited to questionnaires by large vehicle manufactures on specific systems, such as the end-user experience of using ADAS [22]. Such quantitative research has not been used to develop QoE evaluation models, but instead used to improve the vehicle's software directly. Thus, as vehicles are becoming autonomous, the current state of the art does not understand the implication of, for instance, full self-driving on overall performance and end-users' QoE [6]. In addition, it is not clear why does a human alter or how it can affect a self-driving AV. Conducting subjective QoE tests for AV would be complicated and expensive approach, due to the many uncertainties in, for example, vehicle structure, road conditions, type of ride, driver's behaviours, and multi-user interactions. Thus, intelligent behavioral monitoring is necessary to understand the human-robot interaction [21]; but also to evaluate the knowledge created by the autonomous software, which is the scope of the Quality of Context (QoC) layer, discussed in the Section VI-B3.

As a summary, the proposed input and output metrics from the Application layer are the following:

- *Input options* = { *Application type (video, voice, IoT, web); data type and format (e.g. multimedia content); source-node and end-node; packet-header and bitstream information; etc.* }
- *Output metrics* = { *Application metrics (e.g. bit-error rate, encoding, bitrate, content genre); Subjective human scores (e.g. MOS, mood, enjoyment, emotions, easy-of-use, end-user's characteristics and background), Objective human scores (e.g. HVS)* }

*4) **Virtual Layer**:* In general, the virtual layer is a combination of entities that enable virtual data storage and processing. AVs take full advantage of the virtual layer in getting periodic software updates of the AI system. This is typically achieved through the concept of federated learning, in order to preserve customer's data privacy, data security, and data access rights [59]. Therein, each vehicle re-trains its AI model with its own vehicle's data and shares the model with a centralized virtual server which synchronizes the models from the vehicles to generate a unified version of it, which is again shared with the vehicles as a software update. However, the AVs manufactures benefit from analyzing the data produced by the customer's vehicles, for example in learning and developing new services [6], and this is becoming a burden for the virtual layer as a single AV is forecasted to generate 4TB of data daily [3]. Thus, the performance of the platforms for data storage, retrieval, and processing are important in ensuring quality in AVs [60]. Caching is one technology explored in connected cars by making the vehicles distributed caching nodes through V2V links [60], while other studies deploy in-network caching to support the data exchange [34]. Technologies such as cloud and edge computing are also utilized to enable storing and processing of data in virtual decentralized nodes [14].

The massive amount of generated data per vehicle [3] would also lead to rapidly increasing wireless communication demands to ensure reliable data transfer and avoid congestion in the existing networks [10]. Novel technologies emerge in fulfilling the multi-tenant and service-oriented requirements, such as SDN, NFV, and Self-Organazing Network (SON). Such technologies alleviate the burden of the base-stations and cellular core networks by distributing network operations across virtual instances. For example, a virtual network slice is assigned to a specific service with dynamically allocated network resources according to the service requirements [14]. However, enabling intelligent network operations also open up the possibilities of failures, in terms of spectrum allocation, interference, and data management, especially in a mobile AVs environment with frequent topology changes [10]. Thus, analytical techniques are necessary to monitor the performance of the virtual instances, where the vehicle can benchmark the network slice performance [60].

As a summary, the proposed input and output metrics from the Virtual layer are the following:

- *Input options* = { *Specification of available storage, caching nodes and virtual instances; Available computational power and memory for data processing; etc.* }
- *Output metrics* = { *Estimated data storing, processing, and retrieving time; estimated performance of network slices; etc.* }

B. Measure QoIoT

The data acquired from the layered model enables transforming low-level metrics into context-aware analytics by adding ML and knowledge-driven approaches. Measure QoIoT essentially extracts the contextual data analysis process from the object perception layers. As proposed with the QoIoT definition in Section IV, this entity aims to evaluate the quality of an AV from the human and machine perspective. The main goal is to propose KPIs that are understandable for both humans and machines, improving their SA and enhance decision-making processes. To do that, Measure QoIoT proposes three sub-layers that consist of context-aware pattern recognition, novelty, and anomaly detection techniques to evaluate the data generated by the model. Quality of Data (QoD) mainly addresses the metrics generated by the Physical layer, with the aim to test the accuracy of the hardware and software, but also to find anomalies and discover novel states. Quality of Network (QoN) facilitates SA projection of the way networks impact the overall service performance through enhanced network monitoring features, orchestration of virtual functions, prediction algorithms, and pattern recognition strategies. Quality of Context (QoC) processes the metrics from the previous layers and groups them in greater context - QoE and QoME. Finally, QoC discusses the formation of the objective knowledge that evaluates QoIoT.

*1) **Quality of Data (QoD)**:* This layer consists of real-time processing of the generated data in the system, evaluating its accuracy and integrity. To do that, QoD requires the metrics from the Physical layer, but also high-level knowledge of the case-study domain (e.g. the nature of generated data). Each data instance can be described with one or multiple variables, and instances can be related to each other. For example, the nature of the data can be sequential (e.g., time-series), spatial (related to its neighboring instance), spatio-temporal (a combination of sequential and spatial), and graph (interconnected vertices in a graph) [39]. Depending on the data nature, QoD can classify the patterns within data sets and infer proper techniques for conditional monitoring, fault and anomaly detection. For instance, QoD can look for: (1) Point anomalies, considering an individual data instance; (2) Contextual anomalies, when a data instance is anomalous in a specific context, but not otherwise; and (3) Collective anomalies, if a collection of related data instances are anomalous with respect to the entire data set [39]. As noticed from the four-layered QoIoT model in Section IV-A, vehicular data is unique due to the abundance of data sources and the possibility of failures. Vehicular data can certainly be spatio-temporal, where apart from the time-series nature, data instances are related to their neighboring instances. This means that besides point anomalies, the generated vehicular data needs to be evaluated from contextual and collective point-of-view. Evaluating the data-sets from different contexts is itself a complicated process, and therefore will be covered in the QoC layer.

a) Condition monitoring and anomaly detection: Condition monitoring and anomaly detection in spatio-temporal processes require approaches that are more than just detecting the crossing of predefined thresholds. It requires unsupervised ML techniques to autonomously find relationships in the data [61]. For instance, clustering approaches have been commonly used for conditional monitoring. Abbasi et al. [62] provided a detailed review of the available clustering algorithms for wireless sensor networks. However, as the data becomes high dimensional, the relatively simple clustering approaches reach their limit due to data density and slow response [61]. To cope with this, researchers developed more advanced techniques that perform dimensionality reduction, such as PCA [8]. PCA assumes that features with a low variance provide a small contribution to the final model and, therefore, can be neglected by computing principal components. Clustering and dimensionality reduction has been shown as an effective way to improve the success rate of anomaly detection algorithms [39]. Depending on the availability of labeled data, the anomaly detection algorithms are classified as supervised, unsupervised, and semi-supervised. However, their basic scope is first to learn the normal behavior of the system and then detect undesired situations [63]. Chandola et al. [39] conducted an in-depth review of the anomaly detection algorithms, with a type of techniques and their use in real-life applications. Anomaly detection has been heavily used in AVs cases, most commonly for solving the avoiding collisions problem [64]. For instance, Boukhari et al. [65] developed an anomaly detection method for automotive CAN bus sequential data by using Support Vector Machine (SVM). Van et al. [66] used sensory data describing the vehicle's speed, GPS, and acceleration to detect anomalous behaviors of AVs by using Convolutional Neural Networks (CNN). Deep learning and Recurrent Neural Networks (RNN) are also used for anomaly detection in AV, with Long Short Term Memory (LSTM) emerging as the most common technique due to the time-series nature of the data [33]. Multi-sensor anomaly detection on real-world vehicle engine was proposed by Malhotra et al. [33], where LSTM is used to detect anomalies from predictable, unpredictable, periodic, aperiodic, and quasi-periodic time-series.

b) Moving objects recognition: Detection of moving objects is commonly achieved through image recognition [67], video annotation [68], and scanned 3D models [69, 70]. Identifying objects in the near vicinity creates the necessary contextual knowledge for the vehicles to be able to move safely. However, classifying an object is a difficult task due to the high speeds of the vehicle, but also because of the various range, shape, and velocity of objects that can appear on the road, such as pedestrians, obstacles and other vehicles [70]. Potential errors in the classification, but also in the hardware used for detection, may have devastating consequences [69]. Approaches suggest using multiple sources to avoid errors, such as hardware redundancy, where data from multiple vision sensors is used to create fusion algorithms in detecting the same objects [71]. For instance, Gao et al. [71] deployed five radar, three vision, and one position/attitude sensor to detect and classify moving objects. The discussion of their results suggests that monitoring the error rate over time when predicting the same object may indicate faulty prediction. Deep anomaly detection method for video, tested on MNIST moving dataset, is proposed by Ben et al. [72] that uses convolution LSTMs to reconstruct missing frames, where the reconstruction error is used as an indicator for anomaly. Unsupervised deep learning autoencoders are proven to be effective method by first training a model to reconstruct normal data and subsequently identify anomalies as samples with high reconstruction errors [73]. Sabokrou et al. [74] used fully convolutional neural networks to extract discriminative features of video regions. They modeled a normal event as a Gaussian distribution and labeled a test region that differed from the normal reference model as anomaly. A complete review of the approaches to detect anomalies in videos are presented by Kiran et al. [75].

As a summary, the proposed input and output metrics from the QoD layer are the following:

- *Input options* = { *Raw and labeled data generated from the devices; nature of the generated data;* }
- *Output metrics* = { *Results from statistical analysis on the data; clusters in the data; generated principal components; detected anomalies; detected malfunctioning devices; etc.* }

*2) **Quality of Network (QoN)**:* The scope of the QoN layer is to facilitate SA projection of the overall network performance through enhanced network monitoring features, support the orchestration of virtual functions, prediction algorithms, and pattern recognition strategies. The monitoring performed in the 4-layered model from Section VI-A feeds QoN with a wide range of low-level metrics and events. More specifically, QoN process the feedback from the: (1) Network hardware, if available, such as framing errors, lost signals, and fault detection; (2) Detected delays, jitter and packet losses from the individual service applications; (3) Detected wireless access link characteristics, radio conditions, and environmental noise; (4) In-network measurements, if present; and (5) Virtual instances and their performance. Analysis of such metrics is aimed at reasoning and knowledge acquisition in order to deduce conclusions and mitigation responses regarding potential network failures, traffic flow predictions, channel estimations, resource management, load balancing, and network congestion prediction.

a) Playout buffer: The range of V2X communications, illustrated in Figure 6, impose unique requirements for the networks, such as low throughput and high reliability for V2V, and high throughput for infotainment or remotely-controlling for V2I [31]. Watching a high-resolution video in downlink, or live-streaming from the vehicle's cameras in uplink heavily impacts the network, which can generate a domino effect on the quality of other more critical communications, such as V2V. Consider the remotely-controlled mining vehicle as an example, where the live video has to be aligned with the vehicle's CAN bus (e.g., speed, engine control, transmission, fuel) for the expert driver. Moreover, the steering movements by the expert driver must be executed by the vehicle without any perceivable delay [42]. QoN, in this case, is responsible for monitoring the time-synchronization of the data-transmission and playout buffer management. In cases certain thresholds of delay, jitter, and packet-loss are crossed, QoN should act accordingly in decreasing the video resolution or switching off a camera. Multiple studies suggest models and algorithms for managing the playout buffer of video-streaming while experiencing mobile vehicle environment [76, 77].

b) Enhanced network monitoring: The data generated by the 4-layered model enables QoN to integrate ML-based techniques for enhanced real-time evaluation of the network. Measuring congestion in the network is one example where the standard approaches can be optimized to be applicable for AVs cases. For instance, SCReAM [54] and Coupled CC [55] send bulk data to measure the congestion, generating heavy traffic that creates bottlenecks in the network. Such tools are inefficient for AVs cases due to the dynamic mobile environment with constant changes in the topology and resources, requiring constant congestion measurements. A ML model can be trained to predict the results from the cited congestion methods using non-intrusive features obtainable in real-time, that will not flood the network. Studies have modeled wireless access link characteristics, such as RSSI, RSRP, RSRQ, and SINR, to estimate cell load [78] and channel quality [79]. For vehicular cases, Taherkhani et al. [80] reviewed other solutions for estimating network congestions in crowded intersections. Upon detected congestion, the QoN layer should propose mitigation responses in adjusting the communications parameters, such as transmission power, transmission rates, and contention window sizes [32]. Besides network congestion, the research on network quality in vehicular cases has utilized various deep RL, Markov models, and clustering methods to address other network-related issues, such as (1) Location prediction based scheduling and routing [81]; (2) Load balancing and vertical handoff control [82]; and (3) Virtual resource allocation [83], and (4) Estimating latency during drive tests [47].

c) Role of virtualization in AV: SDN enables a holistic view of the network, especially suitable for the network providers, as it can be used to dynamically manage the network and provision QoE. AI-powered SDN solutions are emerging to overcome the high complexities created by decentralizing network infrastructures, operations, and storage. Duan et al. [84] presented an adaptive vehicle clustering scheme, and also a beam-formed transmission method, improving the bit-error rate and throughput performance. Zheng et al. [83] developed an SDN-enabled delay-optimal dynamic virtualization radio scheduling scheme for a connected car. Cai et al. [85] explore software defined V2V communication, developing an discrete stochastic approximation algorithm for network resource sharing among devices under imperfect network state information (NSI). In terms of achieved throughput, their simulated results show large benefits of utilizing virtualization, with their algorithm reducing the gap between perfect and imperfect NSI. SDN and NFV are shown to leverage mobile edge computing and caching, utilizing RL to obtain the resource allocation policy in applications for smart cities [86].

A comprehensive review of ML techniques applied to SDN is provided by Xie et al. [87]. Supervised methods are used for mobility management, leveraging handovers and interworking among WLANs, LPWANs, LTE/NR, where the vehicle in coordination with the SDN controller decides V2I and V2V offloading [87]. The largest benefit of utilizing virtualization within AVs, and the upcoming 5G/NR in general, is tied to creating and orchestrating virtual slices [88]. Afolabi et al. [88] explore the use cases of slicing within core networks and applications. 3GPP has defined three high level virtual slices and a connected vehicle is expected to utilize each of them simultaneously: (1) Enhanced mobile broadband, for high-throughput video streaming and V2I; (2) Ultra low latency, for V2V critical messages; (3) Massive M2M communications, for the abundance of connected vehicles and devices sharing data [88].

d) Network anomaly detection and security: As mentioned before, degradation in network performance (e.g. imposed delays and packets drops) can lure the AV in taking a wrong decision, due to the dependencies on V2V messages [25]. Therefore, the estimation error in network-based anomaly detection should be robust not only against modeling errors and measurement noises but also against transmission delays and data losses caused by the limited capacity of the communication channel [89]. Network anomalies in AVs cases can also be purposely created for intrusion and malicious attacks on the in-vehicle network [90]. Security researchers have proven the vulnerabilities of in-vehicle networks, as a consequence of being exposed to many V2X connections [91], with the vehicles' CAN Bus as a typical target. Kang et al. [92] implemented a deep learning intrusion detection system (IDS), trained with in-network packets to detect malicious attacks on the CAN Bus. While Wang et al. [91] developed a IDS using hierarchical temporal memory to predict anomalous network data flow. Besides attacks on CAN Bus, Loukas et al. [93] implemented an LSTM framework using data from both cyber and physical processes to detect Denial-of-Service, command injection, and malware.

As a summary, the proposed input and output metrics from the QoN layer are the following:

- *Input options* = { *Access to live network; detected delays, jitter, and packet losses; Detected radio conditions and environmental noise;* }
- *Output metrics* = { *Detected network anomalies and security threats; Mitigation responses on detected network failures; Location based predictions on traffic-flow, Playout buffer, Load balancing; Network resource management and virtual resource orchestration; etc.* }

*3) **Quality of Context (QoC)**:* Measuring and reasoning on individual metrics, produced by the previous layers, would not describe the overall system quality and performance [65]; This is due to the complex interconnections between each of the layers and the uncertainties of how individual point anomalies would degrade the QoIoT. The QoC layer aims to group the individual metrics generated from the QoIoT model, QoD and QoN layers, and evaluate them in a greater context (Fig 7). A context would be any information which helps to determine the situation of an entity [5]. Thus, the context in an autonomous service needs to be evaluated from the perspective of the two main identified entities: human and machine, following the QoIoT definition in Section IV. A study by Fridman et al. [9] describes the importance of making the human aware of the decisions by the AI and its performance, as well as making the AI aware of the human presence and HCI. Therefore, QoC measures the objective KPIs from Section V, which are envisioned to create the necessary context understandable for both humans and machines, improving their SA by linking the human-business and business-machine relationships. In that direction, the objective KPIs, such as productivity, safety, and efficiency, are context situations evaluated with the metrics generated by QoIoT, as context attributes. In the following, we provide a guideline for creating the KPIs, as a basis for measuring the experiences of human and machine perspectives.

a) Evaluating QoME: A study by Li et al. [21] provides framework of creating intelligence tasks that will test different parts of the overall AI system, with a conclusion that not well-defined tasks are usually hard to test. Out study attempts to extend the work by LI et al. [21] in the direction of utilizing the intelligence tests, in combination with the service requirements, to build a labeled data that can be used for benchmarking the overall AI system. That is, we envision the identified input and output data-sources from the previous layers to be used in a data-driven approach to test each

ML model against the objective QoME KPIs, described in the Section V, that guarantees each part of the AI system acts according to the requirements. Within the mining case-study, measuring the QoME KPIs, such as production levels, reliability, and efficiency, can be utilized for labeling the generated vehicle's data, which is essential for triggering the self-optimization methods, that greatly use RL [2].

The complex SoS in AVs would consist of several independent AI problems, such as scene perception, vehicle control, trajectory mappings, and localization [9]. Consider the frames generated from the cameras as an example, which may be relevant for few AI models, such as object detection, collision avoidance, and line-keeping. Each of those models may rely solely on the cameras' input or utilize additional data-sources from the layered-QoIoT model. The idea is to benchmark each AI model, considering its data-sources, with the objective QoME KPIs. We argue that the benchmarking process should also consider evaluating QoD and QoN for each AI model separately; This is due to the expectations that an instances of a data-source, such as camera frames, may be anomalous for one model, but not for the other [39]. For instance, the same camera frame may successfully identify the road lines, while simultaneously failing to identify a pedestrian on the road. Those context anomalies are studied in the literature as novelty detection methods, as they may carry out new important information for the system [39]. For instance, if a novel state is measured to improve the fuel/battery efficiency, then the AI models should be re-trained to incorporate that state in their estimations. Thus, QoD and QoN will identify anomalous behaviour of a data-source, while QoC will study the novelty of the detected anomalies. Some popular methods include Bayesian classifiers for estimating continuous distributions [8] and the combination of Density and Class Probability Estimation for one-class classification [94]. Recently, within AV, methods have been developed for novelty detection in 3D lidar data [95], and visual navigation using deep learning with autoencoding [96]. Zhang et al. [97] used state graphs as a behavior model for safe driving, deriving and self-learning novel states or discarding context anomalies.

b) Evaluating end-user's experience: Having the end-users physically present in the vehicle complicates the overall self-driving due to the many ways in which the autonomy can be affected and altered [9]. The way end-users interact with the vehicle is a result of their subjective perception and previous experiences. Thus, QoC requires initial knowledge about the high-level expected roles of the end-users in order to measure the QoE. For instance, a driver may be physically present in the vehicle or supervising via a live-stream from distance. In the latter case, the QoC layer should prioritize the inputs from the QoN evaluation in ensuring fulfilment of the live-streaming QoE requirements. In addition, the extend of which the end-users can intervene in a self-driving is also a valuable initial knowledge. For instance, end-users in the vehicle can affect and alter the self-driving by voice commands, touch screens, buttons, wheel, pedals, consuming multimedia, or by simply unexpectedly opening doors and windows. Today's self-driving commercial vehicles require full attention of the human driver during a self-drive, to recognize, acknowledge, and be prepared to take control and adapt when the systems fails [9].

Multiple research studies and vehicle manufactures collected driver's qualitative input for analysis [98, 99]; however, this is seen as an expensive way to get end-user's feedback on the subjective QoE, due to the amount of work required and the large margin of error [6]. One part of the research community on AVs suggests the use of sensors to detect, interpret, and predict movements of the human body within context, including hand, arm, head, face, and eye movements that can reveal information about the ongoing activity, but also evaluate the subjective QoE [2, 9, 100]. Fridman et al. [9] collected and analyzed naturalistic data of drivers during semi-autonomous driving. They use deep learning methods for monitoring human behavior by using video cabin cameras and vehicle's data to extract insights on how AI can impact HCIs. Sonntag et al. [100] developed an eye-tracking system for automatic detection of human intentions to support driver's activity and subjective QoE prediction.

Endsley [11] argues that incomplete SA can lure the driver in taking wrong decision by overriding the self-driving. Thus, systems which help end-users in achieving higher SA will certainly result with overall performance improvements [11]. In that direction, we foresee that displaying QoME KPIs to the end-users, describing the overall performance of the self-driving, will improve their SA. In the mining case-study, the expert-driver in real-time can supervise metrics such as production levels, reliability, and efficiency and override the autonomy if the machine does not perform well. Such scenarios are considered as edge cases, where the AI systems fails. Herein, a combination of measured subjective QoE and objective QoME metrics can be used to label the edge case, further improving the performance of the machine.

As a summary, the proposed input and output metrics from the QoC layer are the following:

- *Input options* = { *Access to live network, and raw data; Understanding of the service requirements* }
- *Output metrics* = { *Measuring objective KPIs from human and machine perspective; weighted links for measuring each KPI;* }

C. Predict QoIoT

A prediction on the overall QoIoT, as a combination of QoE and QoME KPIs, such as productivity, efficiency, safety, MOS, and technology acceptance, concludes the proposed concept. Figure 7 depicts the role of the stakeholder in defining the SLAs by requesting thresholds for each of the measured KPIs. The stakeholder also defines the importance of each KPIs by selecting weights of how much they contribute to the overall QoIoT. This is a dynamic process that can change in real-time, as the stakeholder can alter the quality levels of a certain KPI(s), and thus, QoIoT must adjust to such a scenario. For instance, at specific days, the stakeholder might expect higher productivity, while at other times, the priority can be saving more of the operational costs. Adjusting the weights is a common research problem in Deep Learning [8], where Neural Networks methods are flourishing by learning from trials and errors. A straightforward way to predict the overall QoIoT value is to multiply each of the estimated KPIs from the QoC layer with their corresponding weights, referred in the literature as a dot product [8]:

$$\sum_{i=0}^{m} w_i x_i \tag{3}$$

Where $x \in [x_1, x_2, ..., x_m]$ are the defined KPIs, while $w \in [w_1, w_2, ..., w_m]$ are the corresponding weights per link.

The purpose of this layer is to forecast in the near future values of the objective KPIs by utilizing recorded time-series metrics. The idea is to prepare the system and the self-optimizing methods for the upcoming states and conditions. The utilized time-series data can be recordings from the raw data, and the labeled metrics from the antecedent layers. Sotelo et al. [63] reviewed a selection of forecasting methods, which entail techniques such as moving average, smoothing, and RNN (e.g. LSTM). Once data is acquired, a time series of size N, and the forecasting horizon T are taken as input parameters for a preprocessing task. The last T elements are subtracted from the original time series, and the remaining N - T elements are used for training the model and forecasting [63]. Then, the subtracted T elements are kept aside as a validation set. Hyper-parameter calibration typically takes place for every forecasting algorithm, where each individual hyper-parameter can be tested with different values, thus allowing the algorithm to be tested with different calibration coefficient until it reaches satisfactory forecasts [63].

QoIoT, in its core, supports feedback a loop from the perception and monitoring stages to the estimations and forecasting the calculated objective metrics. This is enabled by the structured decoupled layered model, from where a point or contextual anomaly can be traced and isolated back to its origins. Popular approaches in autonomous systems include self-healing techniques [101] that diagnose and recover from detected faults in the system. In QoIoT, a top-down approach would monitor the objective KPIs, such as productivity, efficiency, and safety, and in case of poor indicators, a root cause analysis crawls back through the antecedent layers. In our previous work, we have estimated a few metrics part of QoIoT, such as video QoE MOS and one-way latency, also with the possibility of conducting root-cause analysis on detected poor scores [47, 44]. In those studies, decision trees were adopted as an estimation technique as they offer the possibility to plot and analyze each decision of the trees. That is, understand and mark the range of values for each feature that leads to an estimate below a certain threshold.

VII. Conclusion and Future Work

Direct research on QoE for IoT has been conducted in recent years [1, 17, 5]; however, in this study, we identify a clear lack of definition for QoE in IoT with a broader scope. We envision the future emerging IoT services to be orchestrated purely from software, which defines new relationships with the end-users. A human-machine interaction enables the end-users to supervise, alter, or completely override an autonomous process. A fully autonomous process can also operate without the presence of a human, which disturbs the concept of QoE, as the end-users are no longer perceiving the output of the machines. Instead, an autonomous decision is taken via M2M communication among the predictive models. The present QoE management models are verified with subjective tests, which neither captures or understands the implications of quality degradation caused by intelligent machines. Thus, we assert the need to define Quality of Machine Experience (QoME), which defines a common ground for discussion on factors that influence the quality and performance of autonomous services, regardless of the presence of a human.

In this article, we define the concept of Quality of IoT-experience (QoIoT), which combines QoE and QoME in understanding the human-business and business-machine relationships. Further, we examine a case-study of autonomous mining vehicles in identifying the data-sources on which we build a methodology to evaluate QoIoT. The present goal is to utilize the data-sources in measuring the quality of the produced data and evaluate the network; and then build performance indicators based on the contextual input from the mining stakeholder to merge the human-business and business-machine relationships.

Finally, the future direction of this study is to transition from its conceptual form by collecting large-scale data sets from the mining case-study and feeding it to the QoIoT architecture. We envision placing the architecture in each of the mining vehicles to evaluate the QoIoT in real-time, with the aim of improving the service performance. We hope that the proposal of QoIoT will stimulate more interest in research and development of novel ML approaches within QoD, QoN, and QoC layers. Then, we would suggest QoIoT deployment in various other interdisciplinary domains, such as in health-care (e.g. remotely controlled surgery), smart manufacturing, and mitigating natural disaster services.

Acknowledgment

We would like to thank Niclas Ögren, Per Johansson, Irina Cotanis, Roman Zhohov, and Ulf Marklund from InfoVista AB for greatly improving the manuscript.

References

[1] Q. Wu et al. “Cognitive internet of things: a new paradigm beyond connection”. In: *IEEE Internet of Things Journal* 1.2 (2014), pp. 129–143.

[2] S. Jeschke et al. “Industrial internet of things and cyber manufacturing systems”. In: *Industrial Internet of Things*. Springer, 2017, pp. 3–19.

[3] B. Li et al. “Big data analytics for electric vehicle integration in green smart cities”. In: *IEEE Communications Magazine* 55.11 (2017), pp. 19–25.

[4] M. Varela et al. “From service level agreements (SLA) to experience level agreements (ELA): The challenges of selling QoE to the user”. In: *Communication Workshop (ICCW), 2015 IEEE International Conference on*. IEEE. 2015, pp. 1741–1746.

[5] S. Möller and A. Raake. *Quality of Experience: Advanced Concepts, Applications and Methods*. Springer Publishing Company, Incorporated, 2014. ISBN: 3319026801.

[6] R. Van der Mei et al. “State of the art and research challenges in the area of autonomous control for a reliable internet of services”. In: *Autonomous Control for a Reliable Internet of Services*. Springer, Cham, 2018, pp. 1–22.

[7] D. J. Fagnant et al. “Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations”. In: 77 (2015), pp. 167–181.

[8] I. Goodfellow et al. *Deep Learning*. The MIT Press, 2016. ISBN: 0262035618.
[9] L. Fridman et al. "Mit autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation". In: *arXiv preprint arXiv:1711.06976* (2017).
[10] X. Cheng et al. "Wireless Toward the Era of Intelligent Vehicles". In: *IEEE Internet of Things Journal* 6.1 (2018), pp. 188–202.
[11] M. Endsley et al. *Situation awareness analysis and measurement*. CRC Press, 2000.
[12] K. Mitra et al. "Context-aware QoE modelling, measurement, and prediction in mobile computing systems". In: *IEEE Transactions on Mobile Computing* 14.5 (2015), pp. 920–936.
[13] M. Fiedler et al. "A generic quantitative relationship between quality of experience and quality of service". In: *IEEE Network* 24.2 (2010), pp. 36–41.
[14] L. Skorin-Kapov et al. "A survey of emerging concepts and challenges for qoe management of multimedia services". In: *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 14.2s (2018), p. 29.
[15] S. Baraković et al. "Survey and challenges of QoE management issues in wireless networks". In: *Journal of Computer Networks and Communications* 2013 (2013).
[16] K. Mitra. "Quality of experience measurement, prediction and provisioning in heterogeneous access networks". PhD thesis. Luleå tekniska universitet, 2013.
[17] A. Floris et al. "Managing the quality of experience in the multimedia internet of things: A layered-based approach". In: *Sensors* 16.12 (2016), p. 2057.
[18] L. Ang et al. "Deployment of IoV for Smart Cities: Applications, Architecture, and Challenges". In: *IEEE Access* 7 (2018), pp. 6473–6492.
[19] Cisco. "The Smart and Connected Vehicle and the Internet of Things". In: (2017).
[20] J. Contreras-Castillo et al. "Internet of vehicles: architecture, protocols, and security". In: *IEEE Internet of Things Journal* 5.5 (2017), pp. 3701–3709.
[21] L. Li et al. "Artificial intelligence test: a case study of intelligent vehicles". In: *Artificial Intelligence Review* 50.3 (2018), pp. 441–465.
[22] L. Li et al. "Cognitive cars: A new frontier for ADAS research". In: *IEEE Transactions on Intelligent Transportation Systems* 13.1 (2011), pp. 395–407.
[23] Y. Xing et al. "Driver activity recognition for intelligent vehicles: A deep learning approach". In: *IEEE Transactions on Vehicular Technology* (2019).
[24] L. Li et al. "Parallel testing of vehicle intelligence via virtual-real interaction". In: (2019).
[25] G. Araniti et al. "LTE for vehicular networking: a survey". In: *IEEE communications magazine* 51.5 (2013), pp. 148–157.
[26] A. Rangesh et al. "No blind spots: Full-surround multi-object tracking for autonomous vehicles using cameras & LiDARs". In: *arXiv preprint arXiv:1802.08755* (2018).
[27] B. Li et al. "Vehicle detection from 3d lidar using fully convolutional network". In: *arXiv preprint arXiv:1608.07916* (2016).
[28] H. Fahmy et al. "Vehicle Risk Assessment and Control for Lane-Keeping and Collision Avoidance at Low-Speed and High-Speed Scenarios". In: *IEEE Transactions on Vehicular Technology* 67.6 (2018), pp. 4806–4818.
[29] A. Arcos-Garcıa et al. "Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods". In: *Neural Networks* 99 (2018), pp. 158–165.
[30] 3GPP TS 22.185. "Service requirements for V2X services". In: (2018).
[31] 3GPP TR 38.885. "Study on NR Vehicle-to-Everything (V2X)". In: (2018).
[32] L. Liang et al. "Toward intelligent vehicular networks: A machine learning framework". In: *IEEE Internet of Things Journal* 6.1 (2018), pp. 124–135.
[33] P. Malhotra et al. "LSTM-based encoder-decoder for multi-sensor anomaly detection". In: *arXiv preprint arXiv:1607.00148* (2016).
[34] B. Hu et al. "Vehicle-to-vehicle distributed storage in vehicular networks". In: *2018 IEEE International Conference on Communications (ICC)*. IEEE. 2018, pp. 1–6.
[35] R. Picard et al. "Toward machine emotional intelligence: Analysis of affective physiological state". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 10 (2001), pp. 1175–1191.
[36] M. Goodrich et al. "Human–robot interaction: a survey". In: *Foundations and Trends® in Human–Computer Interaction* 1.3 (2008), pp. 203–275.
[37] A. Adadi et al. "Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI)". In: *IEEE Access* 6 (2018), pp. 52138–52160.
[38] M. Noroozi et al. "Boosting self-supervised learning via knowledge transfer". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 9359–9367.
[39] V. Chandola et al. "Anomaly detection: A survey". In: *ACM computing surveys (CSUR)* 41.3 (2009), p. 15.
[40] D. Minovski et al. "Quality of Experience for the Internet of Things". In: *IEEE IT Professional* (2020).
[41] T. Samsonowa. *Industrial research performance management: key performance indicators in the ICT industry*. Springer Science & Business Media, 2011.
[42] U. Andersson. "Automation and traction control of articulated vehicles". PhD thesis. Luleå tekniska universitet, 2013.
[43] S. Dadhich et al. "Key challenges in automation of earth-moving machines". In: *Automation in Construction* 68 (2016), pp. 212–222.
[44] D. Minovski et al. "Analysis and Estimation of Video QoE in Wireless Cellular Networks using Machine Learning". In: *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. 2019, pp. 1–6.
[45] M. Ghazizadeh et al. "Modelling driver acceptance: From feedback to monitoring and mentoring systems". In: *Driver Acceptance of New Technology*. CRC Press, 2018, pp. 51–70.
[46] K. Nam et al. "Wheel slip control for improving traction-ability and energy efficiency of a personal electric vehicle". In: *Energies* 8.7 (2015), pp. 6820–6840.
[47] R. Zhohov et al. "Real-time Performance Evaluation of LTE for IIoT". In: *2018 IEEE 43rd Conference on Local Computer Networks (LCN)*. IEEE. 2018, pp. 623–631.
[48] T. Liu et al. "A bi-level control for energy efficiency improvement of a hybrid tracked vehicle". In: *IEEE Transactions on Industrial Informatics* 14.4 (2018), pp. 1616–1625.
[49] R. Xiong et al. "Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle". In: *Applied energy* 211 (2018), pp. 538–548.
[50] R. Palin et al. "ISO 26262 safety cases: Compliance and assurance". In: (2011).
[51] S. Kim et al. "The impact of cooperative perception on decision making and planning of autonomous vehicles". In: *IEEE Intelligent Transportation Systems Magazine* 7.3 (2015), pp. 39–50.
[52] ITU-T Y.1540. "Internet protocol data communication service". In: (2016).
[53] ITU-T G.1030. "Estimating end-to-end performance in IP networks for data applications". In: (2014).
[54] I. Johansson. "Self-clocked rate adaptation for conversational video in LTE". In: *Proceedings of the 2014 ACM SIGCOMM workshop on Capacity sharing workshop*. ACM. 2014, pp. 51–56.
[55] M. Welzl et al. "Coupled congestion control for RTP media". In: *Transport* 6 (2013), p. 1.
[56] ITU-T G.107. "The E-model: a computational model for use in transmission planning". In: (2015).
[57] ITU-T G.1070. "Opinion model for video-telephony applications". In: (2018).
[58] ITU-T P.1203. "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport". In: (2017).
[59] Tian Li et al. "Federated learning: Challenges, methods, and future directions". In: *arXiv preprint arXiv:1908.07873* (2019).
[60] S. Shah et al. "5G for vehicular communications". In: *IEEE Communications Magazine* 56.1 (2018), pp. 111–117.
[61] A. Maier et al. "Big Data and Machine Learning for the Smart Factory—Solutions for Condition Monitoring, Diagnosis and Optimization". In: *Industrial Internet of Things*. Springer, 2017, pp. 473–485.
[62] A. Abbasi et al. "A survey on clustering algorithms for wireless sensor networks". In: *Computer communications* 30.14-15 (2007), pp. 2826–2841.
[63] M. Sotelo Monge et al. "Reasoning and knowledge acquisition framework for 5G network analytics". In: *Sensors* 17.10 (2017), p. 2405.
[64] S. Atev et al. "Clustering of vehicle trajectories". In: *IEEE Transactions on Intelligent Transportation Systems* 11.3 (2010), pp. 647–657.

[65] M. Boukhari et al. “Proprioceptive Sensors’ Fault Tolerant Control Strategy for an Autonomous Vehicle”. In: *Sensors* 18.6 (2018), p. 1893.
[66] F. van Wyk et al. “Real-Time Sensor Anomaly Detection and Identification in Automated Vehicles”. In: *IEEE Transactions on Intelligent Transportation Systems* (2019).
[67] Y. Tang et al. “Vehicle detection and recognition for intelligent traffic surveillance system”. In: *Multimedia tools and applications* 76.4 (2017), pp. 5817–5832.
[68] S. Dasiopoulou et al. “A survey of semantic image and video annotation tools”. In: *Knowledge-driven multimedia information extraction and ontology evolution*. Springer, 2011, pp. 196–239.
[69] H. Cho et al. “A multi-sensor fusion system for moving object detection and tracking in urban driving environments”. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2014, pp. 1836–1843.
[70] W. Zhang. “Lidar-based road and road-edge detection”. In: *2010 IEEE Intelligent Vehicles Symposium*. IEEE. 2010, pp. 845–848.
[71] H. Gao et al. “Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment”. In: *IEEE Transactions on Industrial Informatics* 14.9 (2018), pp. 4224–4231.
[72] I. Ben-Ari and R. Shwartz-Ziv. “Attentioned convolutional lstm inpaintingnetwork for anomaly detection in videos”. In: *arXiv preprint arXiv:1811.10228* (2018).
[73] C. Zhou et al. “Anomaly detection with robust deep autoencoders”. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2017, pp. 665–674.
[74] M. Sabokrou et al. “Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes”. In: *Computer Vision and Image Understanding* 172 (2018), pp. 88–97.
[75] B. R. Kiran et al. “An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos”. In: *Journal of Imaging* 4.2 (2018), p. 36.
[76] L. Vigneri et al. “Low Cost Video Streaming through Mobile Edge Caching: Modelling and Optimization”. In: *IEEE Transactions on Mobile Computing* (2018).
[77] J. Qiao et al. “Improving video streaming quality in 5G enabled vehicular networks”. In: *IEEE Wireless Communications* 25.2 (2018), pp. 133–139.
[78] S. Vangala et al. *Load estimation in 3GPP networks*. US Patent 9,185,587. Nov. 2015.
[79] X. Xing et al. “Channel quality prediction based on Bayesian inference in cognitive radio networks”. In: *2013 Proceedings IEEE INFOCOM*. IEEE. 2013, pp. 1465–1473.
[80] N. Taherkhani et al. “Centralized and localized data congestion control strategy for vehicular ad hoc networks using a machine learning clustering algorithm”. In: *IEEE Transactions on Intelligent Transportation Systems* 17.11 (2016), pp. 3275–3285.
[81] L. Yao et al. “V2X routing in a VANET based on the hidden Markov model”. In: *IEEE Transactions on Intelligent Transportation Systems* 19.3 (2017), pp. 889–899.
[82] Z. Li et al. “User association for load balancing in vehicular networks: An online reinforcement learning approach”. In: *IEEE Transactions on Intelligent Transportation Systems* 18.8 (2017), pp. 2217–2228.
[83] Q. Zheng et al. “Delay-optimal virtualized radio resource scheduling in software-defined vehicular networks via stochastic learning”. In: *IEEE Transactions on Vehicular Technology* 65.10 (2016), pp. 7857–7867.
[84] X. Duan et al. “SDN enabled 5G-VANET: Adaptive vehicle clustering and beamformed transmission for aggregated traffic”. In: *IEEE Communications Magazine* 55.7 (2017), pp. 120–127.
[85] Y. Cai et al. “Software-defined device-to-device (D2D) communications in virtual wireless networks with imperfect network state information (NSI)”. In: *IEEE Transactions on Vehicular Technology* 65.9 (2015), pp. 7349–7360.
[86] Y. He et al. “Software-defined networks with mobile edge computing and caching for smart cities: A big data deep reinforcement learning approach”. In: *IEEE Communications Magazine* 55.12 (2017), pp. 31–37.
[87] J. Xie et al. “A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges”. In: *IEEE Communications Surveys & Tutorials* 21.1 (2018), pp. 393–430.
[88] I. Afolabi et al. “Network slicing and softwarization: A survey on principles, enabling technologies, and solutions”. In: *IEEE Communications Surveys & Tutorials* 20.3 (2018), pp. 2429–2453.
[89] Z. Gao et al. “A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches”. In: *IEEE Transactions on Industrial Electronics* 62.6 (2015), pp. 3757–3767.
[90] G. K. Rajbahadur et al. “A survey of anomaly detection for connected vehicle cybersecurity and safety”. In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2018, pp. 421–426.
[91] C. Wang et al. “A distributed anomaly detection system for in-vehicle network using HTM”. In: *IEEE Access* 6 (2018), pp. 9091–9098.
[92] M. Kang et al. “Intrusion detection system using deep neural network for in-vehicle network security”. In: *PloS one* 11.6 (2016), e0155781.
[93] G. Loukas et al. “Cloud-based cyber-physical intrusion detection for vehicles using deep learning”. In: *IEEE Access* 6 (2017), pp. 3491–3508.
[94] K. Hempstalk et al. “One-class classification by combining density and class probability estimation”. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2008, pp. 505–519.
[95] L. Wellhausen et al. “Reliable real-time change detection and mapping for 3D LiDARs”. In: *2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*. IEEE. 2017, pp. 81–87.
[96] G. Richter et al. “Safe visual navigation via deep learning and novelty detection”. In: (2017).
[97] M. Zhang et al. “SafeDrive: online driving anomaly detection from large-scale vehicle data”. In: *IEEE Transactions on Industrial Informatics* 13.4 (2017), pp. 2087–2096.
[98] I. Pettersson et al. “Setting the stage for autonomous cars: a pilot study of future autonomous driving experiences”. In: *IET intelligent transport systems* 9.7 (2015), pp. 694–701.
[99] L. Buckley et al. “A qualitative examination of drivers’ responses to partially automated vehicles”. In: *Transportation research part F: traffic psychology and behaviour* 56 (2018), pp. 167–175.
[100] D. Sonntag et al. “Overview of the CPS for smart factories project: deep learning, knowledge acquisition, anomaly detection and intelligent user interfaces”. In: *Industrial internet of things*. Springer, 2017, pp. 487–504.
[101] A. Asghar et al. “Self-Healing in Emerging Cellular Networks: Review, Challenges, and Research Directions”. In: *IEEE Communications Surveys & Tutorials* 20.3 (2018), pp. 1682–1709.

Paper C

Real-time Performance Evaluation of LTE for IIoT

Authors:
Roman Zhohov, Dimitar Minovski, Per Johansson, Karl Andersson

Reformatted version of paper submitted to:
IEEE 43rd Conference on Local Computer Networks (LCN)

2018 IEEE 43rd Conference on Local Computer Networks (LCN)

Real-time Performance Evaluation of LTE for IIoT

Roman Zhohov*, Dimitar Minovski*†, Per Johansson†, Karl Andersson*
* Luleå University of Technology
† InfoVista Sweden AB
Skellefteå, Sweden
Email: *name.surname@ltu.se, †name.surname@infovista.com

***Abstract*—Industrial Internet of Things (IIoT) is claimed to be a global booster technology for economic development. IIoT brings bulky use-cases with a simple goal of enabling automation, autonomation or just plain digitalization of industrial processes. The abundance of interconnected IoT and CPS generate additional burden on the telecommunication networks, imposing number of challenges to satisfy the key performance requirements. In particular, the QoS metrics related to real-time data exchange for critical machine-to-machine type communication. This paper analyzes a real-world example of IIoT from a QoS perspective, such as remotely operated underground mining vehicle. As part of the performance evaluation, a software tool is developed for estimating the absolute, one-way delay in end-to-end transmissions. The measured metric is passed to a machine learning model for one-way delay prediction based on LTE RAN measurements using a commercially available cutting-edge software tool. The achieved results prove the possibility to predict the delay figures using machine learning model with a coefficient of determination up to 90%.**

***Keywords*—IIoT, LTE, QoS, delay, jitter, real-time, critical IoT**

I. Introduction

The fourth industrial revolution is predicted a priori and manifested as Industry 4.0. Smart Factories, Industrial Internet of Things (IIoT) and Cyber-Physical Systems (CPS) are main enabling components of Industry 4.0 which augment the industrial application scenarios and automation [1]. According to the predictions, implementation of IIoT will have a tremendous effect on the global economy [2]. PwCs 2016 Global Industry 4.0 survey respondents expect to see $421 billion in cost reductions and $493 billion in increased annual revenues p.a. [2].

The IIoT highly rely on the existing network infrastructure which enables the creation of private-networks transforming the entire manufacturing process into a smart environment [1]. Initially, conventional telecom networks could not cope with industry-specific requirements for reliable, predictable and efficient communication [3]. Industrial networks were mainly based on diverse deterministic bus technologies to satisfy strict requirements of hard real-time automation systems. IIoT technology offers a vast number of use-cases in the domain of healthcare, logistics, industrial production, supervisory control, robotics, etc. [4–6]. Several of such industrial services are engaging closely to the human lifestyle and privacy, to the extent that a life-threating situation might occur if the delivered quality is not on the desired level. Thus, there is a growing necessity in providing and assuring a real-time exchange of information to guarantee safe operations.

On the other hand, recent advances in communication technologies, especially wireless solutions, and interconnection of numerous embedded systems, creating CPS, result in a convergence of the physical and virtual worlds [1]. It is evident that mobile communications will be a key enabler for IIoT, as a certain degree of QoS may be assured [7]. The study done by Ericsson forecasts around 29 billion connected devices by 2022, of which close to 18 billion will be related to IoT [8]. As the conventional LTE cannot cope with this forecast, one of the requirements for the upcoming 5G standard is supporting massive IoT devices [9]. Besides, 5G technology also addresses critical Machine Type Communication (MTC), which is defined as critical IoT [10]. However, in the context of this paper the definition of critical MTC is not limited to the life-threatening situation, but also includes the risks of interrupting industrial operation, causing significant losses for the business. The typical examples of critical IIoT include tele-remote vehicles, remote surgery, robotics, industrial automation, and control [10]. Such IIoT services pose strict quality requirements on the QoS parameters, such as delay, jitter and packet losses. Those QoS requirements are often addressed in the literature as ultra-reliable low-latency communications (URLLC) [7]. Nokia predicts that reliability and latency requirements will play a vital role in critical IoT communication [11]. For example, autonomous vehicles might require end-to-end latency to be less than 10 ms with block error rate (BLER) down to 10^{-6} [12]. As a result, evaluation and prediction of end-to-end latency of the underlying mobile network is a challenging task that becomes essential for both network provider and industrial stakeholder.

This paper offers a real-world case study of mission-critical IIoT that explores the potential of mine digitalization. For this purpose, a software tool was developed to continuously stream a typical sensor data over the LTE network. The absolute, one-way delay is captured per transmission and further coupled and analyzed with LTE RAN measurements. The idea is to explore the possibility of predicting the end-to-end latency that might occur in real-time transmission. The prediction, as a regression problem, is based on a machine learning model, which builds a knowledge base from the real-time radio measurements.

The rest of the paper is organized in the following order: Section II presents an industrial case-study, identifies types of traffic and the overall systems architecture; Section III reviews related work and discusses the importance of real-time sensor stream; Section IV proposes a tool for evaluation and

978-1-5386-4413-3/18/$31.00

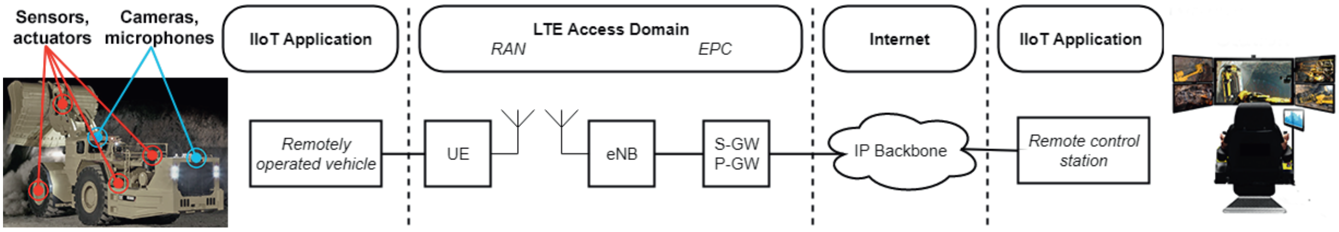


Fig. 1: Systems architecture

prediction real-time performance of LTE; Section V discusses results and application of the tool; Section VI concludes the paper.

II. CASE STUDY

A. Tele-remote mining vehicles

The underground mine is a hazardous environment with a risk of being injured. Moreover, work under these conditions can cause immediate (acute) or long-term (latency) health effects [13]. The motivation of this work is to examine the implementation of tele-remote operation of mining vehicles that will reduce the need for a human operator in the harsh environment of the underground mine.

From the industrial point of view, by introducing the IIoT mining industry may benefit from real-time monitoring, analytics, and control. Mining operations have a set of well-defined metrics such as uptime, productivity, fuel efficiency, etc. For instance, productivity is one important key performance indicator (KPI), defined as an amount of loaded material per hour. Real-time monitoring and data analytics can bring new insights into the best operation strategy based on productivity or fuel efficiency measurements [14].

B. Service components and types of traffic

Industrial use-case of remotely controlled vehicles offers a variety of traffic types and patterns including critical real-time data. Moreover, tele-remote operation and control have attracted significant attention of researchers in recent years [15–17]. The architecture of the system is shown in the Figure 1 and composes remote control station, remotely operated mining vehicle, and communication network. A typical control station for remotely operating vehicles includes:

- Displays showing the surrounding environment, typically including video streams from multiple cameras providing a certain field of view (FOV);
- Speakers providing audio feedback for better context awareness;
- Sensor view displaying metrics regarding vehicle operation and the conditions in the mine;
- Control devices, such as joysticks, wheels, pedals, etc., generating a control stream.
- Health status of the vehicle and the output of monitoring systems;
- Map that shows the machines position to provide situation awareness and facilitate navigation [16];

As a result, the utilized traffic in the system includes:

- Video stream from multiple cameras in uplink (UL);
- Audio stream from multiple microphones in UL;
- Sensor stream in UL from the monitoring systems, motor's controller, and various other externally deployed sensors capturing vehicle's operation and its health status;
- Control stream from the remote control station back to the vehicle in downlink (DL);

The scope of this study is limited to investigating how LTE RAN and radio metrics impact the QoS. The idea is first to examine how the experienced radio and LTE conditions at the end-user, in this case mining vehicles, affect the UL delay. Therefore, an emphasis is given on examining the sensor stream. The findings from this study might give insights in dealing with the complexity of measuring audio and video quality based on jitter, delays and packet losses. Thus, exploring the audio and video streams are left for future work.

III. REAL-TIME SENSOR STREAM

Sensor stream can be further split into two types: critical and non-critical stream. This is an important distinction in the requirements on reliability and end-to-end latency. The purpose of non-critical sensor stream is to periodically send information regarding machine's operation and surroundings. Typically, this data stream does not carry critical information for the machine operation or safety. Whilst, critical real-time sensor stream is sensitive to delays and packet losses since it might negatively affect the business and safety. In ideal case, the delay and jitter for this type of streaming should be upper-bounded following the principles of hard real-time system [18]. Moreover, a real-time sensor stream does not require significant throughput due to small sizes of data chunks [9]. Hence, it creates additional QoS requirements for the network.

Typical examples of critical real-time sensor data in remotely controlled mining vehicles include various vehicle and operation specific parameters, such as: time-to-collision, speed, positioning, motor's metrics, pressure and load on the fork-lift, etc. For instance, latency induced by the network and radio conditions might produce faulty measurement from the speed and proximity sensors, estimating time-to-collision. In addition, delayed sensor stream combined with bad video quality may delude the expert driver and compel a collision. Maximum working speed of the underground mining vehicles can vary up to 15-20 km/h. Simple calculations show that for 20 ms one-way delay and relatively low speed of 11 km/h equals to approximately 6 cm of the displacement which can be critical for this industrial scenario. Thus, one may conclude that QoS degradation on the sensor stream may impact the overall QoE of the IIoT service.

Real-time sensor stream plays a significant role in industrial applications:

- Defines real-time strictness of the scenario and fundamental limitations on the communication technology, protocols, and network architecture;
- Defines the resolution and accuracy of the sensors, and impacts the overall QoE;
- Expands the user-interaction models and enriches the user experience by complementing the video and audio streams. For instance, reduced FOV, degraded depth perception, and image quality result in an inability to estimate speed, time-to-collision, perception of objects, locations and distance to obstacles, and the start of a sharp curve [19]. In such cases, the expert driver will heavily rely on the real-time sensor stream;
- Offers real-time monitoring of industrial processes, enabling to define novel KPIs regarding productivity, efficiency, safety, and reliability;

Studies on remote manipulative control strategies started in the 60s. Authors of [20] show how operators strategy changes with the time delay. Normally, when communication latency is about 1s, the drivers strategy changes to 'move and wait', instead of continuous control. In [21] is demonstrated that movement times increased by 64% and error rates increased by 214% when latency was increased from 8.3 to 225ms. Other studies [16], [17], [22], [23] show that varying delay largely degrades the driving performance compared to constant delay even with a higher magnitude. The unpredictability of time lag can cause over-actuation, such as repeating control commands and over-steering [16]. Performance of LTE network for remote driving was evaluated in [17] by testing the possibility of tele-remote operation of a vehicle under the state-of-the-art commercial LTE network conditions. However, authors have not discussed the evaluation of critical sensor communication.

IV. REAL-TIME PERFORMANCE EVALUATION AND PREDICTION

A. *Background and system's architecture*

Real-time performance of the underlying communication infrastructure is an integral part of QoS in industrial scenarios, due to the critical nature of the services. Evaluation of real-time sensor stream is a challenging task due to the absence of standards and recommendations. The QoS classes recommended by ITU-T in Y-1541 [24] does not address the identified IoT challenges on the communication infrastructure. Moreover, the ongoing work items within ITU addressing data transmission quality techniques, such as G.OM_HEVC, P.NATS, and G.vidmos [25–27] are not intended to assess critical real-time IIoT service. The cited recommendations are targeting multimedia systems, which poses different quality requirements in comparison to the IIoT. Also, utilizing round-trip time (RTT) measurements might not be the most suitable techniques for IIoT due to the abundance of installed sensors and actuators.

To tackle this challenge, a Real-time Tool (RTOOL) was designed and developed as part of this study. RTOOL may periodically send a sensor stream from a source-node to a server (end-node), with a purpose to compute the one-way delay per transmission. Moreover, during the transmissions, the source-node collects radio and RAN measurements in real-time and further use these metrics to find the correlation with the one-way delay in post-processing analysis. The primary goal of RTOOL is to be able to predict the one-way delay in real-time at the source-node based on the collected radio and RAN logs. This prediction utilizes some of the most commonly used machine learning (ML) algorithms and evaluates performance in terms of coefficient of determination and Mean Absolute Error (MAE). Predicted figures for one-way delay could be used to calculate latency budgets for critical IoT, raise alarms, introduce possibilities to reduce latency and perform root-cause analysis.

Figure 1 depicts a high-level view of the utilized communication network. The IIoT case-study consists of remotely operated mining vehicle connected to the terminal (UE) in LTE RAN. Evolved packet system (EPS), which is composed of LTE RAN and EPC, forms the IIoT access network. The packet data network gateway (P-GW) provides connectivity to the public IP network, linking the mining vehicle with the remote control station. The mine where the experiments were carried on is located in the north of Sweden, with fully deployed LTE coverage inside. Radio dots from Ericsson [28] are used as small indoor cells, connected to the nearest outdoor eNB. From the eNBs, the UL traffic goes to the local core, located in Stockholm, from where enters the public Internet to find the end-node also hosted in Sweden but on a different network. The network infrastructure in the mine is designed to have a couple of handovers from one eNB to another. Hence, the Radio Dots are grouped and connected to different outdoor eNBs, for splitting the load.

Each element of the communication system introduces a varying delay due to the connectivity procedures, scheduling, and network fluctuations. Authors of [29] provide a detailed

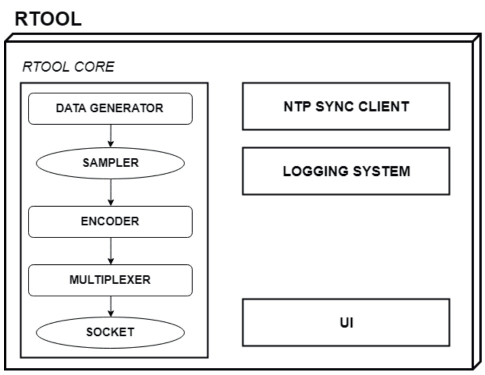


Fig. 2: RTOOL architecture

overview of the latencies that can occur in LTE access domain. Connection establishment on the control and user plane is the most significant part of the latency in LTE access domain and can take up to 106 ms. Moreover, additional delays can be introduced because of scheduling, retransmission, and processing on user plane (up to 28 ms). However, these figures assume that the radio coverage is ideal and the quality of the signal is not degraded. In this paper, the aim is to analyze the delay which occurs due to the degraded radio conditions in LTE RAN. Evaluation of the control plane requires access to the core-network and analysis on metrics such as routing diagnostics, queuing length, load, bandwidth utilization, etc. Such quality metrics give insights to the performance of the network, but licensed RANs are typically complex and hardly accessible [30]. This means that performance evaluation of the control plane requires post-processing, offline analysis on the gathered core-network logs. Therefore, the latency induced by the control plane is out of scope for this study and the focus of the evaluation is on the user data plane. The main reason is the real-time access to the radio measurements at the source-node, such as Received-Signal-Strength-Indicator (RSSI), throughput, Signal-to-Interference+Noise-Ratio (SINR), etc. The hypothesis under test is a real-time analysis of the radio metrics and RAN events for predicting network performance parameters, such as the absolute delay, jitter and packet losses.

Delay figures of the IP backbone will vary depending on the region, network load and the number of hopes between P-GW and application server. For instance, the delay can vary from 15 ms up to 150 ms in Europe [29]. In this work, we assume that application server (e.g., remote control station) is placed close to the P-GW and this delay is negligible.

B. Experiment setup

RTOOL is intended to mimic a sensor stream sent from real tele-operated mining vehicle. Logical components of the designed tool are illustrated in Figure 2. The software tool consists of following components: RTOOL Core, Adaptive NTP client for synchronization, logging system for post-processing and simple User Interface (UI). RTOOL Core has few configurable parameters for each logical element:

- Data Generator mimics a real sensor and generates sensor data with specific size, type, and format;
- A sampler which specifies the sampling rate of the sensor stream based on application requirements;
- Encoder which encrypts the sensor data and formats it according to the requirements;
- Multiplexer combines the data from several sensors into one stream;
- Socket is used to send sensor data from the phone to the server using a specified protocol, UDP by default;

End-to-end or OTT latency measurements refer to the time it takes to send a packet from the source-node until it is received at the end-node. These measurements are done by periodically sending UDP packets with a sensors payload from RTOOL to the end-node. The experiment is performed using different time periods (sampling rate) to send the data. Each UDP packet is time-stamped at RTOOL for seamless analysis at the end-node. Time-stamps and network measurements from TEMS Pocket are stored in log files for offline analysis.

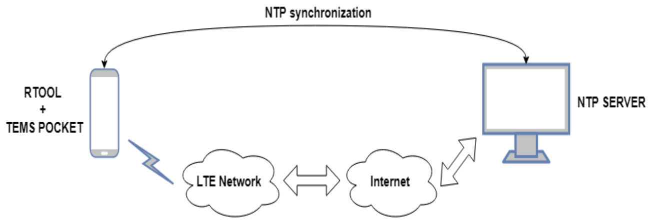


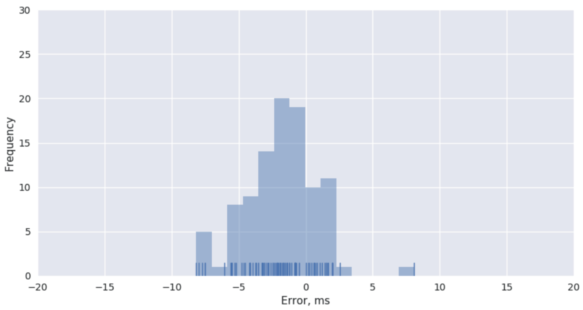
Fig. 3: NTP synchronization setup

RTOOL is envisioned to be part of a previously developed, commercially available software - TEMS Pocket [31]. It is described as a state-of-the-art phone-based test tool developed for measuring the performance and quality parameters of wireless networks. The main functionality of TEMS Pocket is the following:

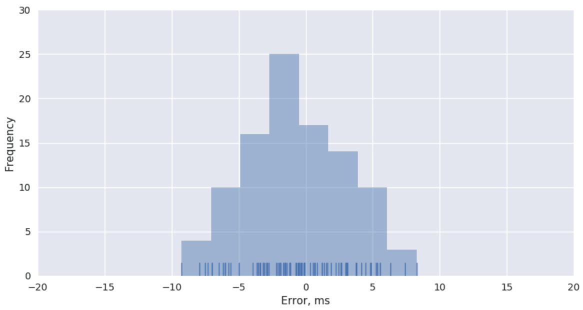
- A real-time radio measurements and event data collection;
- Event data collection from the RAN;
- Indoor and outdoor testing of wireless networks;
- Drive testing capabilities with positioning;
- Storing the radio and event logs for post-processing using other TEMS-ecosystem tools such as TEMS Discovery [32];

Using TEMS Pocket limits the scope of implementation options which means that RTOOL must be implemented on commercial mobile phone or tablet under Android OS. Supported devices by TEMS Pocket are: Sony, HTC, LG, and Samsung. TEMS Pocket supports the following mobile technologies: LTE, WCDMA/HSDPA/HSUPA, GSM/GPRS/EDGE, and CDMA/EV-DO, with a possibility to lock specific radio access technology (RAT) and band. Moreover, TEMS Pocket has several control functions to modify the devices behavior in the LTE network. Control functions work in real-time and allow to perform quick and non-intrusive tests.

In a real-life scenario, it is envisioned a mobile phone to be physically placed on the mining vehicles and be remotely accessible. The user may start recording the radio conditions and RAN events, which will automatically calculate a predicted value of the delay in real-time. In the performed experiments to gather data for building the ML predictions, a Samsung Galaxy S9 mobile phone was equipped with RTOOL and TEMS Pocket. An overview of the experimental setup is depicted in the Figure 3. TEMS Pocket was locked to LTE network and was able to capture more than 1500 parameters explaining the radio and LTE RAN conditions. TEMS Pocket was configured to record those parameters every 5ms during the sensor streaming. RTOOL was set to periodically stream data on different intervals, every 20, 50, 100, and 200ms.



(a) Mobile operator A



(b) Mobile operator B

Fig. 4: Synchronization error

C. *Time synchronization*

Evaluation of real-time performance requires precise estimation of end-to-end delay between the control center and tele-remote vehicle. This task can be achieved by having two nodes perfectly synchronized with each other. Synchronization of the devices in the network is a complex task that can be done in two main ways:

- Synchronization over the network, using various protocols and services;
- Synchronization using external clock references, using signals from Global Navigation Satellite System (GNSS), atomic clocks, etc.;

Conventional approach is to synchronize two nodes by utilizing external references, such as GNSS signals from Global Positioning System (GPS), GLONASS, GALILEO or COMPASS. It is possible to provide accurate time synchronization typically better than 100 nanoseconds to UTC [33]. However, due to the inability to receive GNSS signals in underground environments, such a technique is not suitable for the mining industry. Thus, synchronization for our case-study may be only achieved using the existing LTE network. Main protocols that were developed to keep nodes over the network synchronized are Network Time Protocol (NTP) and Precision Time Protocol (PTP) known as IEEE Standard 1588-2008 [34].

NTP is de-facto time-keeping standard across the Internet [35]. NTP organizes clocks in a layered hierarchal way in terms of a stratum. The stratum level specifies the distance between the reference clock and time server which is used for synchronization. The accuracy of the synchronization which might be achieved using NTP is less than 1ms in LAN and 10ms over WAN [35]. Compensation of the clock's offset is performed by measuring RTT to NTP server. The crucial assumption that NTP makes at this step is that the link is symmetrical and in ideal case UL and DL delays are equal.

For the purpose of the described case-study, a modified version of NTP is developed. Slightly changed topology is used since absolute synchronization to UTC time is not needed as long as NTP server and RTOOL are synchronized to each other. In conventional NTP topology, nodes synchronized to the NTP servers will have different clock errors due to the network fluctuations and clock offsets on the reference NTP servers. The proposed solution implements stand-alone NTP server and have its clock as a reference for the entire system. In this case, the error is completely mitigated on one side since synchronization error on the used server is equal to zero. Figure 3 shows proposed topology for experiment using NTP-based synchronization. RTOOL synchronizes phones hardware clock to NTP server adaptively which means that synchronization can be done only under excellent/good radio conditions (Table I).

TABLE I: LTE signal quality

Signal quality	**Radio parameters**		
	RSRP, dBm	***RSRQ, dB***	***CQI***
Excellent	>-90	>-9	>10
Good	-90 ... -105	-9 ... -12	9 ... 7
Fair	-106 ... -120	<-13	6 ... 1
Poor	<-120		0

It is important to outline that LTE wireless link is not symmetrical due to the differences in UL and DL radio technologies, scheduling mechanisms and bandwidth. However, measurements showed that clock error was acceptable for our scenario and provides the best effort that can be achieved in this specific use-case. Clock error was measured on the live network of two mobile operators by connecting phone directly to the NTP server using Android Debug Bridge (ADB). The results are shown in Figure 4.

D. *Latency prediction*

In recent years, ML models were successfully used in various applications from bioinformatics to speech and image recognition [36]. ML tries to construct data-driven models that can capture complex and sometimes hidden dependencies. ML is becoming increasingly useful with the recent developments of hardware (GPU and TPU), software (TensorFlow and Scikit-Learn) and distributed data-processing frameworks (Hadoop and Spark) [37–39].

The task of real-time network performance prediction from the radio measurements perfectly fits into the ML approach. Authors of [40] provide a general workflow (Figure 5) that can be used to build the ML model for predicting the one-way delay.

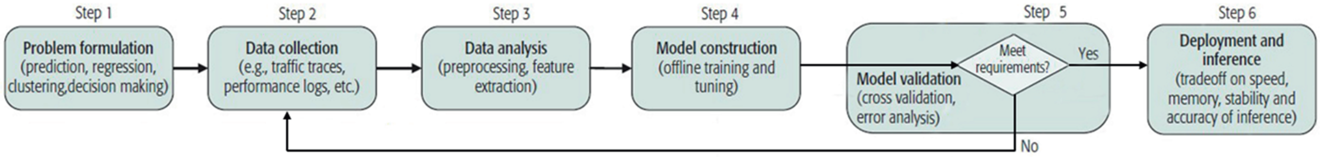


Fig. 5: The typical workflow of machine learning for networking [40]

D.1) Problem formulation and Data collection

The first step of the workflow is a problem formulation. As was stated earlier the goal is to predict network delay caused by the radio environment in LTE RAN. For the second step, as explained in Section IV-B, RTOOL and TEMS Pocket will be used to collect timestamps, radio measurements, RAN events and also send an IP sensor stream. Hence, there are two datasets, from RTOOL and TEMS Pocket.

D.2) Data processing and Feature extraction

As stated earlier, the latency is generated by several factors, but only a few of them (i.e., features) have the most effect on the target metric. The goal of the third step is to pre-process the data by cleaning, formatting and performing feature engineering. Captured radio metrics and end-to-end delay figures are presented as time-series data. End-to-end delay measurements are presented as a series of timestamps that were collected at the source-node and end-node. The time difference between two consecutive timestamps is defined by sampling period which can vary for different experiments. Missing data handling is one of the first steps of the data cleaning process. Commonly recommended ways for fixing missing data are [41]:

- Discarding observations that have missing value;
- Filling the missing values based on other data points, when appropriate;

In this work, both methods were utilized to pre-process measurements from TEMS Pocket. Observations that had less than two unique values were dropped from the dataset. The next step is the datasets alignment, as the TEMS Pocket measurements do not have the same timestamps and frequencies as the delay measurements. The collection of radio and RAN data happens every 5ms, while the delay measurements depend on the chosen sampling frequency. Therefore, the result RAN measurements were mapped to delay timestamps with Piecewise Aggregate Approximation (PAA) [42]. This means taking the median of the TEMS Pocket measurements for the duration of each transmission period of the sensor stream.

Feature engineering is the major step of the entire process of ML model creation. Better features utilization enables simpler models and produce improved results. Raw measurements should be transformed in certain ways to get better results. Considering the nature of radio and RAN measurements, there are several ways to create new features from time-series data:

- Lag features that represent measured values of the captured metrics at prior time samples.
- Window features that represent a value obtained from the values over a fixed window of prior time. Similarly to lag features, this type of features are obtained by taking the mean value of measurements over the fixed window of prior time;

In practice, lag features are obtained by shifting original time series measurements. In general, the gathered data may succumb to stochastic or deterministic time series patterns of single or multiple seasonality, trends and cycles, which typically generate biased predictions [43]. Regarding the mining case-study, the presence of various trends and cycles are obvious, such as lunch breaks and shift changes. Also, the radio conditions inside the mine do not drastically change over time, as it is an isolated and constant environment. Thus, creating additional repetitive cycles. However, the monitored parameters heavily depend on the workflow of RAN, EPC and Public Internet. For instance, the recorded data may be different during the same experienced cycles due to heavy loaded or unloaded eNBs, EPC or Public Internet. This demands additional seasonal analysis on the mentioned entities, which is a complex topic that requires further research. For this reason, the computed lagged features are essential for capturing the relation between latency and RAN conditions that were before transmission.

Collected data have various features with values in different ranges. Most of the ML algorithms are sensitive to features scaling. All features and label (i.e. delay) were scaled using Standard scaler [44]. The scaled feature x' of original feature is given by:

$$x' = \frac{x - \overline{x}}{\sigma(x)}$$

where $\overline{x}$ is the mean value for x, and $\sigma(x)$ is the deviation.

As a result of feature engineering, a whole dataset consists of timestamps, delay measurements from RTOOL, original radio and RAN measurements from TEMS pocket, lag and window features that were obtained from the raw measurements. TEMS Pocket may collect more than 1500 parameters (i.e., features), but for latency prediction basic L1 radio measurements were utilized. These measurements are extensively described in following 3GPP standards [45–47] and detailed description of LTE L1 measurements is beyond the scope of this paper. Parameters such as RSRP, RSRQ, RSSI, SINR and physical throughput were considered as a features for the model construction. Main motivation for such feature selection is to enable less intrusive prediction and examine the possibility of predicting latency from L1 measurements from UE. Finally, a Pearson correlation coefficient was computed among all the recorded features [48]. From the results, the one-way delay correlates the most with the physical throughput, and RSSI.

D.3) Model construction

The goal of the model construction step is to select an appropriate ML algorithm to get a reliable model with the best prediction. In this paper, we evaluated different ML-regressors from scikit library [49], with the most consistent validation results achieved from:

- Artificial Neural Networks (MLP);
- Decision Tree Regressor [50];
- Model ensembling: Bagging technique with a Decision Tree Regressor as a base for bagging ensembling.

Bootstrap aggregation, or bagging, is a technique that can be utilized with many classification and regression algorithms to reduce the variance associated with prediction, and as a result provide higher accuracy. Original dataset is divided into many bootstrap samples, after that base method is applied to each bootstrap sample and then the predictions are combined, by averaging for regression, to obtain the overall result, with smaller variance [51].

D.4) Model validation

The next step after model construction is its validation which is considered to be an important phase that verifies the model's accuracy and ensures that it does not overfit. Separate experiments on two live commercial LTE networks were performed to obtain datasets for training and validation. After all the measurements, all regressors were trained using the training set (80%) and evaluated against the testing set (20%).

Models were assessed by evaluating the accuracy of the delay prediction. Results for each sampling period are shown in Table II. More detailed discussion of the results follows in the next section. However, from the Table II it is obvious a linear decrease in the prediction accuracy as the sampling rate increases. Performance varies with a sampling rate due to the number of radio measurements and events collected within a time series. Lower sampling periods allow to monitor network with higher resolution and capture all fluctuations of the radio environment.

TABLE II: Prediction performance

	R^2 **(coefficient of determination)**			**Mean Absolute Error (MAE)**		
Sampling period	NN (MLP)	Decision Tree	Bagging Decision Tree	NN (MLP)	Decision Tree	Bagging Decision Tree
20 ms	82 %	82.2 %	90.7 %	0.23	0.11	0.091
50 ms	75.5 %	77.7 %	85.1 %	0.28	0.16	0.15
100 ms	73.7 %	67.3 %	81.8 %	0.29	0.19	0.13
200 ms	60 %	50 %	66.8 %	0.35	0.31	0.22

V. Results and Discussion

Splitting the data-set into 80/20 for training and validation would mean predicting the one-way delay on 20% of the data-set. As it is difficult to find blind spots and lousy radio environment on a live commercial LTE network, the "silent" box from the Figure 6(a) was used for generating training data. This radio frequency (RF) shield degrades signal quality or completely blocks it. Thus, it can produce a full range of radio conditions. A mobile phone will experience massive radio signal fluctuations when it is placed inside, while constantly opening and closing the box. Therefore, the TEMS Pocket data was recorded in this manner. The Figures 7(a) and 7(b) depicts massive fluctuations with the delay when the mobile phone streams sensor data while experiencing lousy radio conditions. At some point, the delay jumps all the way to 2000ms. The blue line on the Figure 7 is the true recorded delay, while the orange line is the predicted delay.

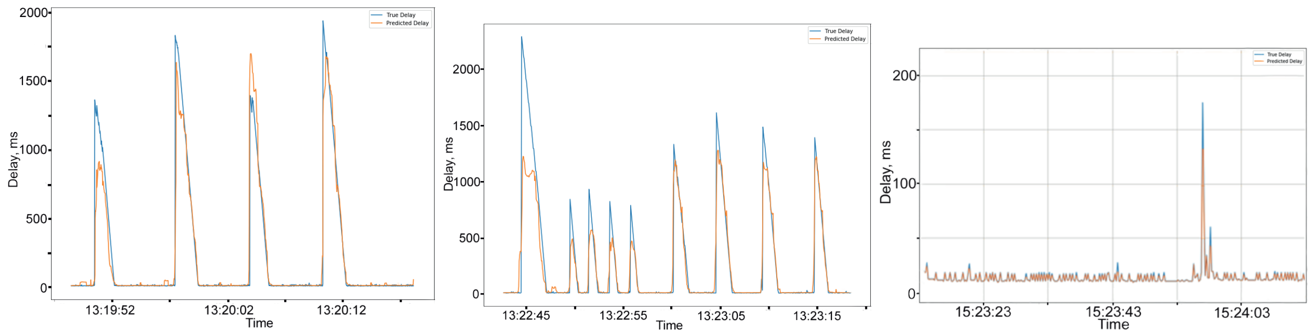
(a) RF shield (b) Drive test in the mine

Fig. 6: Experiments

Real-Life underground drive tests as on Figure 6(b) were also performed to validate the prediction accuracies. Figure 7(c) illustrates the experienced delay while streaming from the mine. The values are relatively low and constant, with one recorded peak of 200ms during a handover. However, the prediction model was able to capture the handover and successfully predict the occurring delay. Again, the blue line is the true recorded delay, while the orange line is the predicted value. The benefit from such software implementation, prediction model and results is the ability to compute in real-time the one-way delay. A mobile phone with TEMS Pocket can be physically placed on the mining vehicle that will measure radio conditions and RAN events in real-time. These measurements are fed to the ML model which computes a prediction of the one-way delay. The predicted values may be plotted on one of the displays in the remote control room as a real-time gauge chart, to give better context awareness for the expert driver.

VI. Conclusion

The research community and the industry acceptance of IoT suggest rapid digitalization of industrial processes. Being applied in various domains, each IIoT service requires prioritization of different KPIs and service requirements. Thus, the network performance evaluation becomes linearly more complex as each IIoT requires different QoS assurances. In this work, a real-world industrial scenario was analyzed to evaluate the importance of critical real-time sensor streaming. For this purpose, a software tool was developed to capture the absolute, one-way delay for each transmission. The latency metrics



(a) RF shield, sampling period 20 ms (b) RF shield, sampling period 50 ms (c) Drive test in the mine

Fig. 7: Comparison of true delay and predicted delay figures

and further analyzed with various LTE RAN measurements. A machine learning technique is used to grasp the relation between the latency metrics and the captured radio measurements. The contribution of this study is a delay prediction for each transmission in real-time based on the correlation and learning processes. The initial results prove the possibility to estimate delay figures caused by the LTE RAN events and radio disturbances from the environment. The highest accuracy of the prediction is estimated at 90%.

The approach taken in this study is the first step in assessing the performance of an IIoT service. The achieved results enable further calculation of latency budgets for a given critical IoT service, as well as opens the possibilities to reduce latency and perform root-cause analysis.

Acknowledgment

This research work is financially supported by the Erasmus Mundus program PERCCOM [52], facilitated by the European Commission. The authors are grateful to the InfoVista Sweden AB that supported this work and provided access to TEMS products and research facilities. A special thanks to Niclas Ögren, network & protocols specialist, and Ulf Marklund, R&D manager, for helpful discussions and support.

References

[1] H. Kagermann, W. Wahlster, and J. Helbig, "Recommendations for implementing the strategic initiative industrie 4.0: Final report of the industrie 4.0 working group," April 2013.
[2] PWC, "Industry 4.0: Building the digital enterprise."
[3] M. Wollschlaeger, T. Sauter, and J. Jasperneite, "The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0," *IEEE Industrial Electronics Magazine*, vol. 11, pp. 17–27, March 2017.
[4] J. Marescaux, J. Leroy, M. Gagner, F. Rubino, D. Mutter, M. Vix, S. E. Butner, and M. K. Smith, "Transatlantic robot-assisted telesurgery," *Nature*, vol. 413, no. 6854, p. 379, 2001.
[5] A. Sajid, H. Abbas, and K. Saleem, "Cloud-assisted IoT-based SCADA systems security: A review of the state of the art and future challenges," *IEEE Access*, vol. 4, pp. 1375–1384, 2016.
[6] L. Sanchez, L. Muñoz, J. A. Galache, P. Sotres, J. R. Santana, V. Gutierrez, R. Ramdhany, A. Gluhak, S. Krco, E. Theodoridis, *et al.*, "Smart-Santander: IoT experimentation over a smart city testbed," *Computer Networks*, vol. 61, pp. 217–238, 2014.
[7] Ericsson, "Manufacturing reengineered: robots, 5G and the industrial IoT," *Ericsson Business Review*, no. 4.
[8] Ericsson, "Ericsson mobility report," 2017.
[9] A. Osseiran, J. F. Monserrat, and P. Marsch, *5G Mobile and Wireless Communications Technology*. Cambridge University Press, 2016.
[10] Q. Zhang and F. H. P. Fitzek, "Mission critical IoT communication in 5G," in *Future Access Enablers of Ubiquitous and Intelligent Infrastructures*, 2015.
[11] Nokia, "White paper: 5G for mission critical communication achieve ultra-reliability and virtual zero latency," tech. rep., Nokia Solutions and Networks Oy, Espoo, Finland, 2016.
[12] "METIS D1.1: scenarios, requirements and KPIs for 5G mobile and wireless system," tech. rep., 2013.
[13] E. L. Petsonk, C. Rose, and R. Cohen, "Coal mine dust lung disease. New lessons from an old exposure," *American journal of respiratory and critical care medicine*, vol. 187, no. 11, pp. 1178–1185, 2013.
[14] B. Frank, L. Skogh, and M. Alakla, "On increasing fuel efficiency by operator assistance systems in a wheel loader," in *Proceedings of International Conference on Advanced Vehicle Technologies and Integration*, July 2012.
[15] C. Zhou, N. Damiano, B. Whisner, and M. Reyes, "Industrial internet of things: applications in underground coal mines," in *Mining engineering*, vol. 69, pp. 50–56, December 2017.
[16] J. Y. C. Chen, E. C. Haas, and M. J. Barnes, "Human performance issues and user interface design for teleoperated robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, pp. 1231–1245, Nov 2007.
[17] R. Liu, D. Kwak, S. Devarakonda, K. Bekris, and L. Iftode, "Investigating remote driving over the LTE network," in *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '17, (New York, NY, USA), pp. 264–269, ACM, 2017.
[18] P. A. Laplante and S. J. Ovaska, *Real-Time Systems Design and Analysis: Tools for the Practitioner*. Wiley-IEEE Press, 4th ed., 2011.
[19] J. Bao, Z. C. Wu, J. J. Wu, P. P. Pei, and C. Liu, "The Impact of Image Parameters of Indirect Viewing Systems on Driver Fatigue and Task Performance," in *Advanced Information and Computer Technology in Engineering and Manufacturing, Environmental Engineering*, vol. 765 of *Advanced Materials Research*, pp. 2046–2050, Trans Tech Publications, 10 2013.
[20] T. B. Sheridan and W. R. Ferrell, "Remote manipulative control with transmission delay," *IEEE Transactions on Human Factors in Electronics*, vol. HFE-4, pp. 25–29, Sept 1963.
[21] I. S. MacKenzie and C. Ware, "Lag as a determinant of human performance in interactive systems," in *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, (New York, NY, USA), pp. 488–493, ACM, 1993.
[22] B. Watson, N. Walker, W. Ribarsky, and V. Spaulding, "Effects of variation in system responsiveness on user performance in virtual environments," *Human Factors*, vol. 40, no. 3, pp. 403–414, 1998. PMID: 9849102.
[23] J. C. Lane, C. R. Carignan, B. R. Sullivan, D. L. Akin, T. Hunt, and R. Cohen, "Effects of time delay on telerobotic control of neutral buoyancy vehicles," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, vol. 3, pp. 2874–2879, 2002.

[24] ITU, "Recommendation Y.1541 Network performance objectives for IP-based services," recommendation, Telecommunication Standardization Sector (ITU-T), 12 2011.
[25] ITU, "Recommendation P.1203.1: Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport - Video quality estimation module," recommendation, Telecommunication Standardization Sector (ITU-T), 10 2017.
[26] ITU, "Recommendation P.913 : Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment," recommendation, Telecommunication Standardization Sector (ITU-T), 03 2016.
[27] ITU, "Recommendation G.1071 : Opinion model for network planning of video and audio streaming applications," recommendation, Telecommunication Standardization Sector (ITU-T), 11 2016.
[28] Ericsson, "Ericsson Radio Dot System, https://www.ericsson.com/ourportfolio/radio-system/radio-dot-system."
[29] N. Nikaein and S. Krea, "Latency for Real-Time Machine-to-Machine Communication in LTE-Based System Architecture," in *17th European Wireless 2011 - Sustainable Wireless Technologies*, pp. 1–6, April 2011.
[30] 3GPP, "Feasibility study for Further Advancements for E-UTRA (LTE-Advanced)," Technical Report (TR) 36.912, 3rd Generation Partnership Project (3GPP), 03 2017. Version 14.0.0.
[31] InfoVista, "TEMS Pocket, https://www.infovista.com/tems/products-for-radio-and-core-networks/radio-network-engineering/portable-testing-for-wireless-networks."
[32] InfoVista, "TEMS Discovery: Discover actionable insights into network performance, https://www.infovista.com/tems/products-for-radio-and-core-networks/radio-network-engineering/ran-data-analysis-optimization."
[33] Endace, "IEEE 1588 PTP clock synchronization over a WAN backbone," whitepaper, Endace Technology Limited), 2016.
[34] IEEE, "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," *IEEE Std 1588-2008 (Revision of IEEE Std 1588-2002)*, pp. 1–300, July 2008.
[35] P. Corcoran, "A Matter of Timing: Consumer Electronics and Network Time," *IEEE Consumer Electronics Magazine*, vol. 2, pp. 20–25, Oct 2013.
[36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[37] P. Sherkhane and D. Vora, "Survey of deep learning software tools," in *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*, pp. 236–238, Feb 2017.
[38] A. Parvat, J. Chavan, S. Kadam, S. Dev, and V. Pathak, "A survey of deep-learning frameworks," in *2017 International Conference on Inventive Systems and Control (ICISC)*, pp. 1–7, Jan 2017.
[39] V. Sze, Y. H. Chen, J. Emer, A. Suleiman, and Z. Zhang, "Hardware for machine learning: Challenges and opportunities," in *2018 IEEE Custom Integrated Circuits Conference (CICC)*, pp. 1–8, April 2018.
[40] M. Wang, Y. Cui, X. Wang, S. Xiao, and J. Jiang, "Machine learning for networking: Workflow, advances and opportunities," *IEEE Network*, vol. 32, pp. 92–99, March 2018.
[41] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research, Cambridge University Press, 2006.
[42] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and information Systems*, vol. 3, no. 3, pp. 263–286, 2001.
[43] S. F. Crone and N. Kourentzes, "Feature selection for time series prediction–a combined filter and wrapper approach for neural networks," *Neurocomputing*, vol. 73, no. 10-12, pp. 1923–1936, 2010.
[44] scikit learn, "sklearn.preprocessing.StandardScaler, http://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html."
[45] 3GPP, "Physical layer procedures (FDD)," Technical Specification (TS) 25.214, 3rd Generation Partnership Project (3GPP), 03 2010. Version 8.9.0.
[46] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation," Technical Specification (TS) 36.211, 3rd Generation Partnership Project (3GPP), 03 2017. Version 14.2.0.
[47] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," Technical Specification (TS) 36.213, 3rd Generation Partnership Project (3GPP), 03 2017. Version 14.2.0.
[48] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, Inc., 1 ed., 1997.
[49] scikit learn, "Scikit-learn: Machine Learning in Python, http://scikit-learn.org/stable/."
[50] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, pp. 81–106, Mar. 1986.
[51] C. D. Sutton, "11 - classification and regression trees, bagging, and boosting," in *Data Mining and Data Visualization* (C. Rao, E. Wegman, and J. Solka, eds.), vol. 24 of *Handbook of Statistics*, pp. 303 – 329, Elsevier, 2005.
[52] J. Porras, A. Seffah, E. Rondeau, K. Andersson, and A. Klimova, "Perccom: A master program in pervasive computing and communications for sustainable development," in *2016 IEEE 29th International Conference on Software Engineering Education and Training (CSEET)*, pp. 204–212, April 2016.

Paper D

Analysis and Estimation of Video QoE in Wireless Cellular Networks using Machine Learning

Authors:
Dimitar Minovski, Chrsiter Åhlund, Karan Mitra, Per Johansson

Reformatted version of paper submitted to:
IEEE 11th Internetional Conference on Quality of Multimedia Experience (QoMEX)

2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)

Analysis and Estimation of Video QoE in Wireless Cellular Networks using Machine Learning

Dimitar Minovski*†, Christer Åhlund*, Karan Mitra*, Per Johansson†
* Luleå University of Technology
† InfoVista Sweden AB
Email: *name.surname@ltu.se, †name.surname@infovista.com

***Abstract*—The use of video streaming services are increasing in the cellular networks, inferring a need to monitor video quality to meet users' Quality of Experience (QoE). The so-called no-reference (NR) models for estimating video quality metrics mainly rely on packet-header and bitstream information. However, there are situations where the availability of such information is limited due to tighten security and encryption, which necessitates exploration of alternative parameters for conducting video QoE assessment. In this study we collect real-live in-smartphone measurements describing the radio link of the LTE connection while streaming reference videos in uplink. The radio measurements include metrics such as RSSI, RSRP, RSRQ, and CINR. We then use these radio metrics to train a Random Forrest machine learning model against calculated video quality metrics from the reference videos. The aim is to estimate the Mean Opinion Score (MOS), PSNR, Frame delay, Frame skips, and Blurriness. Our result show 94% classification accuracy, and 85% model accuracy (R^2 value) when predicting the MOS using regression. Correspondingly, we achieve 89%, 84%, 85%, and 82% classification accuracy when predicting PSNR, Frame delay, Frame Skips, and Blurriness respectively. Further, we achieve 81%, 77%, 79%, and 75% model accuracy (R^2 value) regarding the same parameters using regression.**

***Keywords*—QoE, QoS, Video, MOS, PSNR, LTE**

I. Introduction

The proliferation of video streaming services goes hand-in-hand with the users' trend of communicating through video messages, causing large spikes in the projections of video-traffic. Cisco forecasts 396 exabytes (EB) IP traffic per month by 2022 (122 EB in 2017), out of which 82% will be video traffic [1]. Moreover, forecasting 77.5 EB of mobile data traffic per month by 2022, out of which 78% will be video [1].

The research community has been focused on video quality assessment (VQA) of video-streaming services, breaking down the metrics impacting the Quality of Service (QoS) and end-user experiences [2–4]. Therein, a great portion of the prior work is adopting the ITU definition of Quality of Experience (QoE) [5]. Guidelines and standards were proposed which describe the steps of conducting evaluation of the quality perceived by the end-user [6]. However, the complexity of commercial VQA rises, as service-level agreements (SLAs) are signed among service provides, operators, and customers, to promise certain quality levels. Depending on the service requirements, the VQA models are classified as [2]: full reference (FR), reduced reference (RR), and no-reference (NR). Our focus are the NR models, as unlike the others they do not require post-processing video analysis and are closest to a real-time VQA.

978-1-5386-8212-8/19/$31.00

A review of the most prominent NR models suggest relying on metrics retrieved from the video bit-stream and packet-header, such as bitrate, packet loss rate (PLR), video encoding, etc. [7]. Therein, there are growing concerns on the translation from the mentioned metrics into end-user QoE, captured through Mean Opinion Scores (MOS) [8]. Moreover, the most popular video streaming services are starting to largely encrypt the header and payload information, giving even less chance for the NR models to measure the traditional QoS/QoE metrics. Thus, standards developed by ITU Study Group 12, such as P.NAMS and P.NBAMS [9] fall short due to encryption and various real-time video services, such as Video over LTE (ViLTE) and Over-the-top (OTT). This becomes even more evident in a dynamic and mobile environment, where packets are dropped due to poor coverage, interference, cell edges, and handovers, in cases of real-time streaming.

In this study we argue for scope extension of the NR based methods, by utilizing collection of objective log data describing the radio conditions of an LTE link (e.g. poor coverage, interference, cell edge, etc.). For this purpose, experiments were conducted on commercial live LTE network, in a city and underground mine environment, with both stationary and drive tests. A total of 200 video files were streamed via Real-time Transport Protocol (RTP) in uplink, from a mobile phone to a video server. During the real-time streaming process the mobile phone measures metrics that describe the radio LTE link, such as RSSI, RSRP, RSRQ, CINR, etc.

Our contribution is an in-depth analysis on the dependencies and relationships among the radio and video quality metrics. Further, the aim is to find the root-causes for a quality degradation of a real-time streaming video content by analyzing the radio metrics at the source-node.

A machine learning (ML) model is developed to predict conventional video quality metrics solely by using the measured radio metrics. Herein, metrics such as MOS, PSNR, frame delay, frame skips, and blurriness were computed on the streamed video files in post-processing by using ITU-T J.247 (PEVQ) [10]. The idea is to use PEVQ's metrics as a reference point to train towards a ML model with the recorded radio logs. The results on average show 87% and 79% of model accuracy (R^2) with classification and regression respectively when predicting the video quality metrics.

The rest of the paper is organized as follows: Section II present the related work; Section III describes the case-study and the experimental setup; Section IV shows the achieved results; Section V concludes the study and discusses future work.

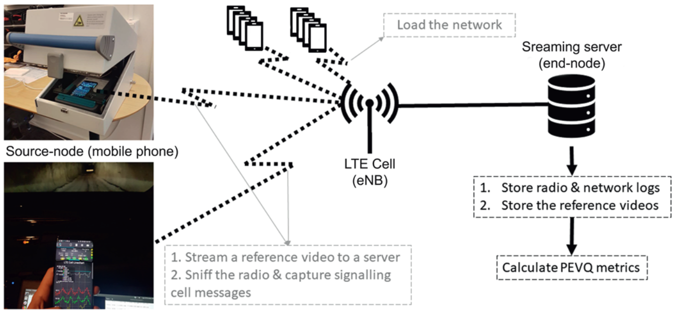


Fig. 1: The general infrastructure of the experiments.

II. Related work

Multiple survey studies on VQA classify the existing models as FR, RR, and NR [2], [7]. In this respect, ITU established standardization bodies to produce recommendations and guidelines for VQA, among which is ITU Study Group 12. The models referred as FR require post-processing analysis on the original with the degraded video reference file. Studies which explore FR models typically compare the corresponding frames of the two video files and produce objective quality metrics such as Bluriness, Frame Freezes, Frame Skips, PSNR, etc. For instance, PEVQ [10] computes such metrics and models the behavior of the human visual tract to calculate a MOS from 1 to 5. NR based models have gained attention for their ability to perform quality monitoring in real-time. Their classification is further explained in [7] as: pixel-based (NR-P), bitstream (NP-B), and hybrid models combining the first two. The NR-P use a decoded representation of the transmitted signal and analyze the quality based on the pixel information. Thus, NR-P models require high computational complexity which makes them unfit for real-time VQA [11], especially in mobile streaming services. NR-B based models have their own classification explained in [7]. One of the first works in the parametric VQA models [12], [13] explore the impact of bitrate, framerate, and PLR on the video quality, without taking the video content into account. More advanced NR-B ML-based models emerged [14–17], where additional parameters are considered from the bitstream and codec settings to estimate MOS.

In [14] the work describe extracting metrics from the packet-header and uses encoder/decoder information to build a model for predicting video quality of IPTV service. However, as pointed out by [18], [19], the existing solutions relying on packet inspection are for the most part no longer viable, mainly due to encryption. Going over the encryption issues, ITU-T with the P.1203 recommendation [20] defined modes depending on the encryption levels. In [16] the work describe the use of ITU-T P.1203 to build various ML models for MOS prediction on Youtube videos. The study uses in-network features mainly based on calculating mathematical statistics on lists that store the amount of data downloaded by the client in 1 or 5 second periods. ITU-T P.1203 and thus [16] are tailored for HTTP adaptive streaming, targeting issues such as re-constructing missing frames, or buffering of downlink traffic. However, we consider real-time video services, such as ViLTE, IPTV, Skype, surveillance cameras, or other OTT services, where the uplink stream suffers from different issues, mainly frame/packet dropping.

Feature	Description	Mode
RSSI	*Carrier Received Signal Strength Indicator: comprises the linear average of the total received power (in W) observed only in OFDM reference symbols [21]*	Idle
RSRP	*Reference Signal Received Power: the linear average over the power contributions (in W) of the resource elements that carry cell-specific reference signals [21]*	Idle
RSRQ	*Reference Signal Received Quality: the ratio N*RSRP/(Carrier RSSI), where N is the number of resource blocks of the carrier RSSI measurement bandwidth [21]*	Idle
SSS Cell Power	*Secondary Synchronization Signal power for detected cells (in dB) [22]*	Idle
Total RS Power	*Total Reference Signal power calculated from serving cell RSRP and channel bandwidth [22]*	Idle
CQI	*Channel Quality Indicator for code word 0 [22]*	Conn.
PUSCH MSC	*Modulation Coding Scheme index for the uplink transport block [21]*	Conn.
RS CINR	*Carrier to Interference plus Noise Ratio of the signal carrier best servings for the intervention seemed at all other sites/sectors, plus all the noise*	Conn.

TABLE I: Collected parameters at the source-node

In [15] the work describe the use of in-smartphone measurements to predict a subjective QoE of different applications, including Youtube, Facebook, and Google Maps. Even though the scope of this study is also on downlink traffic, the authors use RSSI as part of their features in predicting a MOS. Based on their results, RSSI is strongly correlated to the Youtube MOS predictions, and therefore serves us as motivation to conduct more in-depth VQA solely on radio metrics. In [23] the authors point out that besides poor coverage major QoS deficiencies may cause various other uncontrollable radio and cellular network phenomena, such as interference, handovers, cell edges, and congested cells. The authors of [24] analyze the effect of RSSI, RSRP, SINR, and RSRQ on events such as throughput, link adaptation, packet scheduling, and handover. In [25], the fluctuations in the experienced radio conditions were modeled to predict one-way latency. However, limited number of studies discuss mapping the radio metrics to QoE, especially in multimedia context. We identify the study [17] as a comprehensive work to examine in-smartphone radio measurements to predict QoE metrics, however, targeting web-browsing and voice applications.

III. Experimental Setup

A. High-level cellular experimental setup

The general infrastructure used to conduct the experiments is depicted on the Figure 1, with LTE as an underlying cellular network. A mobile phone resides as a source-node, streaming the videos to a server. The experiments were performed in: 1) city environment, with under 50,000 inhabitants; 2) Underground mining environment, with completely isolated LTE network; The reason for choosing two different environments are two-fold. First, the proposed ML model should understand the different interference factors and radio conditions in such two disparate environments; Secondly, an option to access the cell logs from the mobile operator in the mine; Having isolated LTE network in the mine, with no interferences, enables the ML model to understand the way congested cells impact the video quality. Heavy loaded cell will create packet drops and

	RSRP (dBm)	SSS Cell Power	RSRQ (dB)	CQI Code	PUSCH MCS	RS CINR (dB)	Total RS Power (dBm)	RSSI (dBm)	MOS	FrameDelay	PSNR	FrameSkips	Blurriness
count	200	200	200	200	200	200	200	200	200	200	200	200	200
mean	-88.34	17.43	-8.29	12.55	17.78	22.03	-65.31	-60.28	3.77	-88.17	39.26	1.66	1.73
std	13.87	0.54	1.84	2.30	5.83	7.89	14.21	13.51	1.10	43.49	4.26	2.79	0.32
min	-117.06	14.81	-12.61	6.07	1.99	5.75	-93.61	-88.45	1.18	-277.57	22.64	0.00	1.24
max	-61.47	17.85	-4.81	15.00	23.12	30.00	-38.01	-35.31	4.81	69.55	42.19	16.94	3.53

Fig. 2: Statistics from the combined data-set.

force HARQ and RLC re-transmissions [22]. Hence, progressive load was generated during the video streaming in the mine and the cell logs were used to confirm the load. To generate the load, 15 phones were equipped with a modified TWAMP application [26], with each phone streaming continuous bursts of synchronized 1500 bytes UDP packet trains in UL and DL, in a coordinated manner to avoid sending data simultaneously. Each of the 15 phones on average streamed data with 70 Mbps. After analyzing the mining cell logs, the videos were streamed with cell load in ranges of: $<5\%$, 10-20%, 35-50%, 50-65%, 75-95%. Finally, the described cell loading technique was used also in the city experiments, due to the relatively small population size and making sure that there is active load on the serving cells. All the experiments were performed with SIM cards from one Swedish cellular operator, without any speed or data limitations.

The source-node had pre-installed commercially available application - TEMS Pocket [27], with capability to measure radio metrics on a millisecond level basis. The measured metrics (Table I) are retrieved by scanning the radio and analyzing the signals from the surrounding cells, part of the LTE standard [21]. Therein, ”Idle” metrics suggest a state when the phone is passively scanning the radio, while ”RRC_Connected” (or ”Conn”) is when the source-node attempts to send user-data over the cell. Due to almost perfect LTE environment in the city, a shield box was used (Figure 1) with the source-node inside in order to capture heterogeneous Table I metrics. Opening and closing the box during the experiments would damper the signal and yield poor radio signal, simulating poor LTE coverage. The statistics from the behaviour of the LTE signal during the experiments can be seen on Table 2. It may be observed that the collected radio conditions range from the worst to the best possible, according to the 3GPP specification [22], suggesting the diversity and reproducibility of the experiments. Generally, the experiments were further divided into: 1) Stationary and 2) Drive tests - where the source-node is naturally communicating with the serving cell; 3) Stationary and 4) Drive tests with shield box - where the LTE signal is dampered; The experiments were equally performed separately on each of the four setups, in both city and underground mine environment. During the drive tests the speed limit inside the mine was 10 km/h, while in the city average speed was around 40 km/h.

B. *Video sequence setup*

The work conducted for this study was part of the PIMM-DMA project [28], aiming to digitize underground mines with capabilities to control the mining vehicles from distance. The vehicles were equipped with 720p HD cameras to live-stream video in uplink, to a remote-control room, while driving in the mine. Therefore, the service requirements defined the video encoding and transmission parameters used in the experiments, shown on the Table II.

Parameter	Value
GOP length	30 frames
Frame rate	30 fps
Bitrate (avg)	3 Mbps
Resolution	HD 720x480p
Color mode	YUV (4:2:0)
Codec	H.264
Protocol	RTP

TABLE II: Video encoding and transmission parameters.

To perform the experiments, Samsung Galaxy S9 was used as a source-node to stream the video files via RTP, utilizing the open-source ffmpeg library [29]. Thus, an ffmpeg script was running on the mobile phone and the video server, for streaming and receiving the video files respectively. The choice of using ffmpeg is due to the robustness of the library, with abilities to change various video transmission and encoding parameters on-the-fly, but also to monitor sending and receiving logs in real-time during streaming. In practice, ten video files were used as samples for the streaming process, each with duration of 8 seconds, as recommended by OPTICOM - the creator of PEVQ. Five video files were recordings from the mine, while the other five were retrieved from the Video Quality Experts Group (VQEG) public database. At the end of the experiments a total of 200 videos were streamed and stored on the video server for further analysis.

IV. The impact of radio parameters on the video quality metrics

The goal of this section is to explore the dependencies and relationships among the cellular radio metrics and the objective video quality metrics. Further, the aim is to find the root-causes for a quality degradation of a real-time streaming video content.

A. *Data gathering process and methodology*

During the real-time streaming process of each video a log file was recorded at the source-node using TEMS Pocket, capturing the radio parameters described in the Table I, later referred as features. As TEMS Pocket measures most of the features on a millisecond level basis, an arithmetical mean had to be computed on each of the features for the duration of each video (8 seconds). Further in post-processing, the original video files were compared to the corresponding streamed videos by using PEVQ [10]. As an output, PEVQ computes various video quality metrics, among which MOS, PSNR, frame delay, frame skips and blurriness. Herein, the PEVQ metrics are the labels, to which the features may be trained against. Few statistics on the gathered data-set can be seen on

	RSRP (dBm)	SSS Cell Power	RSRQ (dB)	CQI Code	PUSCH MCS	RS CINR (dB)	Total RS Power (dBm)	RSSI (dBm)
MOS	0.375	0.402	0.585	0.382	0.360	0.389	0.369	0.309
FrameDelay	0.013	-0.004	0.052	0.007	0.001	0.000	0.028	0.020
PSNR	0.245	0.266	0.451	0.233	0.215	0.218	0.240	0.191
FrameSkips	-0.218	-0.087	-0.307	-0.130	-0.188	-0.117	-0.214	-0.185
Blurriness	-0.468	-0.579	-0.515	-0.559	-0.519	-0.580	-0.467	-0.420

Fig. 3: Pearson's correlation on the features against PEVQ metrics.

the Figure 2. The data-set combines the measurements from all experimental setups described in Section III-A.

Before conducting any ML predictions, a correlation analysis on the gathered data-set may be computed. Pearson's correlation was used to calculate the correlation between each of the features with the corresponding labels (see Figure 3). From the results one may conclude that constantly most correlated parameters are RSRP, RSRQ, and RSSI. This is not surprising as it was already shown that those are the main features in describing the radio link under drive tests [30].

The next step is utilizing the most suitable ML techniques to build a model with training and validation. Our study follows the data-driven QoE analysis approach, proposed in [3], where the use of decision trees for QoE prediction come after a correlation analysis, followed by a QoE causality analysis as a final step. Decision trees were also previously used for similar context in [15], [16] to predict video MOS. Further, as pointed out in [3], decision trees are humanly readable and interpretable, which perfectly matches the scope of our work - understand the trees for enabling a root-cause analysis. Thus, techniques such as Neural Networks, Deep Learning, Support Vector Machine, and Bayesian Networks were not considered as it is challenging to extract and describe their decisions that lead to a estimation [31].

B. Video quality prediction with a classifier

According to the ITIL [32], among other matters, an SLA defines service-level targets which may be measured through key performance indicators. Regarding video streaming applications, these service-level targets are clearly stated in [33], where the most acceptable threshold for MOS score is arguably 3.5. Thus, a classification ML problem may be defined for the described case-study, where all the recorded MOS values below the 3.5 threshold are assigned with 0, and above that with 1. Graphical representation of the recorded data-set with the MOS threshold split is shown on the Figure 4.

	MOS	PSNR	Frame Delay	Frame Skips	Blurriness
R^2	*0.94*	*0.89*	*0.84*	*0.85*	*0.82*
RMSE	*0.192*	*0.236*	*0.317*	*0.298*	*0.359*
		Feature Importance			
RSRP	*0.197*	*0.187*	*0.278*	*0.210*	*0.144*
Cell Power	*0.043*	*0.063*	*0.054*	*0.073*	*0,087*
RSRQ	*0.293*	*0.193*	*0.369*	*0.283*	*0.271*
CQI	*0.123*	*0.083*	*0.064*	*0.013*	*0.034*
MSC	*0.012*	*0.051*	*0.012*	*0.108*	*0.066*
CINR	*0.178*	*0.128*	*0.109*	*0.149*	*0.173*
RS Power	*0.021*	*0.092*	*0.030*	*0.016*	*0.127*
RSSI	*0.133*	*0.176*	*0.174*	*0.148*	*0.098*

TABLE III: Random Forest classifier results.

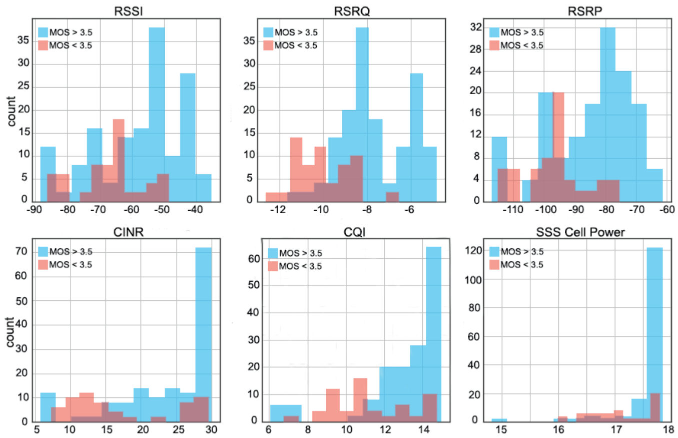


Fig. 4: Data-set split based on MOS, greater or below 3.5.

Therein, one may observe the patterns of the MOS distribution, such as majority of high MOS values tend to appear under high RSRP, RSRQ, CINR, and RSSI values.

Similarly to the MOS threshold of 3.5, by observing the Figure 2, one may define threshold values for the other metrics, such as PSNR of 35 dB, Frame Delay of -84 ms, Frame Skips of 1.5%, and Blurriness of 1.5%. The reasons for choosing such objective video quality metrics are two-folded. First, a model with estimated objective metrics may be applied to various video services, regardless of the end-users and their subjective opinions. Second, commercial video-streaming companies are already looking at objective metrics such as PSNR, delays and packet losses to improve the delivered quality of their service [34], [35].

Random Forest (RF) Classifier was utilized first to train and validate the recorded features against the labels [36]. RF brings more robustness to the predictions due to the random selection of the features for the splits at each node. Other ensemble ML algorithms such as Bootstrapped Aggregation use all the features for each split and hence, in our case-study, produce decision trees with structural similarities. The classifier was tested with various parameters for regularization, known as pruning, and most stable results on average achieved 10 trees with minimum sample split of 10, which limits the depth of each tree in order to avoid over-fitting [37]. The whole data-set of 200 samples were divided with 80/20 split, where 160 samples were used for training, against 40 sample for validation. In practice, 5 RF models were tested, for training and validation the features against MOS, PSNR, Frame Delay, Frame Skips, and Blurriness respectively. Table III shows the results from the classifier, calculating R^2 and $RMSE$ values for the estimation, and also presenting the RF features importance. It is important to note that the results from Table III were achieved after 5-fold cross-validation, which is further discussed in Section IV-E.

C. Conducting a root-cause analysis

The possibility to visualize the RF's trees enables interpretation of the achieved results. This is suitable for conducting a root-cause analysis upon estimation of a poor video quality. LIME is a open-source framework [38] which enables interpretation of complex ML models, in this case by analyzing

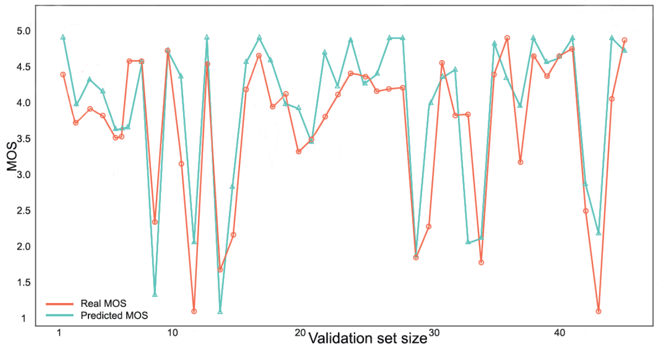


Fig. 5: Results from the Random Forest regressor.

the 10 decision trees. LIME scans each decision of the trees and marks the range of values for each feature that lead to an estimate below or above the threshold. The actual trees are not shown due to lack of space, but an average of their decisions is shown on Table IV. Such table may be generated for any of the labels, and in this case MOS > 3.5 is presented.

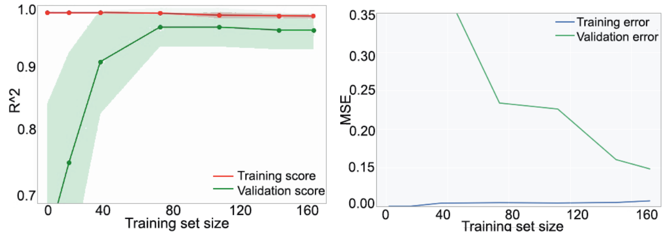
Table IV may be seen precisely as a root-cause analysis, from where few conclusions emerge. For instance, measured RSSI value below -70 dB would advocate bad coverage areas or blind spots, while CINR value below 13 suggests interference and therefore produces poor video quality. Moreover, similar RSRQ values of the serving and neighboring cell indicates cell edge and therefore a handover may be expected, which might cut the video streaming for a short time.

As mentioned before, load from 15 other phones was generated on the same cell while conveying the video traffic. Since the cell logs from the underground mine were available, the precise cell load was observed and correlated to the features and the labels. It has been already shown that RSRP, RSRQ, and CINR may suggest congested cells, resulting with network packet drops [39]. Our results prove this hypothesis, as the MOS scores on average fell from 4.6 to 2.1 as the cell load increased from 30% to 95% in the mine environment. Moreover, the high accuracy percentage from Table III shows that it is possible to capture the way cell load impacts the video quality.

Even through the city and mine seem such disparate environments, the results show they complement each other. That is because the ML model is trained to understand some unique phenomena, which pattern will be captured solely in such isolated environment like the mine. Examples are signal reflections and the way multipath is handled in LTE. Since there is no interference in the mine from end-users and inter-cells, the pattern of the RSRP, RSRQ and CINR are different

Metrics	MOS > 3.5
RSRP	*-98.3 < RSRP > -93.9*
SSS Cell Power	*17.1 < SSS Cell Power > 18.4*
RSRQ	*-9.3 < RSRQ > -8.4*
CQI	*12.2 < CQI > 13.5*
MSC	*18.4 < MSC > 20.2*
CINR	*13.3 < CINR > 14.5*
RS Power	*-76.3 < RS Power > -64.7*
RSSI	*-70.2 < RSSI > -63.9*

TABLE IV: Minimal range of values to achieve MOS > 3.5.



(a) Classifier (b) Regressor

Fig. 6: Learning curves.

compared to the city case. Thus, the pattern of the mining radio metrics describe the reflections from the rocks and how it affects the video quality during streaming. A conclusion is that training the model on more unique cases will even further improve the performance.

D. Regression method against video quality metrics

After high accuracy with RF classifier, we also attempt to construct a regression problem and directly predict the PEVQ metrics using the same features. RF regressor with the same regularization as the classifier was initialized, due to the same reasons as stated before. Figure 5 plots the true computed MOS (red) and estimated MOS (cyan). The model achieved accuracy of 85% (R^2), with $RMSE$ of 0.361. The achieved results for the other metrics are as follows: PSNR R^2 score of 81%, Frame Delay 77%, Frame Skips 79%, and Blurriness 75%, with RMSE values of 0.364, 0.382, 0.379 respectively.

E. Dealing with the over-fitting problem

One common technique to test a ML model from being over or under fitted is a Learning Curve [40]. The idea is to observe the changes of the errors, both in the training and validation, as the training set size increases. Figure 6 plots learning curves for RF (a) classifier and (b) regressor. The size of the training set was set to $\{1, 20\ 35, 75, 110, \text{and } 140\}$. On top of that, a 5-fold cross-validation was performed when selecting the samples, which takes random samples for each of the training size and averages the results. Examining the Figure 6(b), the learning curve first attempts to use 1 sample for the training set and test it against the validation set. In such case, the model fits perfectly the training data, thus training error of 0. However, a model with only one training set will perform poorly on the un-seen validation set, as illustrated. This indicates over-fitting. Further, as the training set size increases, the validation error gradually decreases, reaching an arguably acceptable MSE value of 0.15. Thus, figure 6 also shows the rate of change in both variance and bias. Low training error is associated with low bias, which indicates that the model is also not under-fitted. However, the trend of the validation error in Figure 6(b) indicates that increasing the training set size will yield lower MSE, which is indication to do more experiments.

V. Conclusion and Future work

The current models for conducting real-time VQA are challenged with tighten security and encryption of the video services. Further, the real-time nature of services such as ViLTE, Skype, and OTT, does not fit into the concept of adaptive streaming, since frames are dropped instead of retransmitted or buffered. Hence, this study looks for alternative ways to enrich the performance of the current models for VQA. The important finding is that there is a correlation between the measured quality of the radio LTE link and the video quality metrics. Thus, our approach utilizes Random Forrest to estimate MOS, PSNR, frame delay, frame skips and blurriness, based on radio metrics such as RSSI, RSRP, RSRQ, and CINR. Moreover, we conduct a root-cause analysis by inspecting and describing the decision of the estimations. In real scenario, the Random Forrest classifier and regressor reside at the end-user, where radio metrics are used to estimate video quality metrics. A future work is understanding how the service providers can utilize such estimations to improve their service quality.

Regarding the machine learning model, there are few potential improvements as a future work: 1) Include static parameters such as video resolution, frame-rate, codec, etc as features to generalize the model; 2) Exploring and testing additional radio/network metrics is one aspect, while extracting the existing feature is another. For instance, breaking down a single feature, or grouping multiple features into a set of subfeatures may capture unique real-life cases and hence improve the performance; 3) Training the model on more unique cases, such as subway/train ride, large crowd stadiums, rural areas, etc; 4) One may create sub-ML problems within the model to estimate MOS, such as estimating load on the serving cell or predicting a handover; and 5) One may also attempt to first predict objective metrics, such as PSNR and Frame Delay, and use them as additional features to estimate MOS.

Acknowledgment

This work was conducted under the PIMM-DMA project [28]. We would like to thank Niclas Ögren and Irina Cotanis from InfoVista AB, the reviewers and the shepherd from the IEEE QoMEX conference for improving the quality of this work.

References

[1] Cisco, "Cisco visual networking index: Forecast and trends, 2017 to 2022," *Whitepaper*, 2017.
[2] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE transactions on broadcasting*, vol. 57, no. 2, p. 165, 2011.
[3] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A tutorial on video quality assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.
[4] K. Mitra, A. Zaslavsky, and C. Åhlund, "Context-aware qoe modelling, measurement, and prediction in mobile computing systems," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 920–936, 2015.
[5] ITU, "ITU P.10/G.100 Amendment 1: Definition of QoE," 2006.
[6] ITU, "ITU G.1011: reference guide to Quality of Experience assessment methodologies," 2016.
[7] M. Shahid, A. Rossholm, B. Lövström, and H.-J. Zepernick, "No-reference image and video quality assessment: a classification and review of recent approaches," *EURASIP Journal on image and Video Processing*, vol. 2014, no. 1, p. 40, 2014.
[8] T. Hoßfeld, P. E. Heegaard, M. Varela, and S. Möller, "Qoe beyond the mos: an in-depth look at qoe via better metrics and their relation to mos," *Quality and User Experience*, vol. 1, no. 1, p. 2, 2016.
[9] ITU-T, "SG12: performance, QoS and QoE; Available: goo.gl/wnudy8,"
[10] ITU, "ITU-T Recommendations J.247: Objective perceptual multimedia video quality measurement in the presence of a full reference," 2008.
[11] M. T. Vega, E. Giordano, D. C. Mocanu, D. Tjondronegoro, and A. Liotta, "Cognitive no-reference video quality assessment for mobile streaming services.," in *QoMEX*, pp. 1–6, 2015.
[12] ITU, "ITU-T Recommendations G.1070: opinion model for video-telephony applications," 2007.
[13] A. Raake, M.-N. Garcia, S. Moller, J. Berger, F. Kling, P. List, J. Johann, and C. Heidemann, "Tv-model: Parameter-based prediction of iptv quality," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pp. 1149–1152, IEEE, 2008.
[14] M.-N. Garcia and A. Raake, "Parametric packet-layer video quality model for iptv," in *Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on*, pp. 349–352, IEEE, 2010.
[15] P. Casas, A. D'Alconzo, F. Wamser, M. Seufert, B. Gardlo, A. Schwind, P. Tran-Gia, and R. Schatz, "Predicting qoe in cellular networks using machine learning and in-smartphone measurements," in *Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on*, pp. 1–6, IEEE, 2017.
[16] I. Orsolic, M. Suznjevic, and L. Skorin-Kapov, "Youtube qoe estimation from encrypted traffic: Comparison of test methodologies and machine learning based models," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, IEEE, 2018.
[17] V. Pedras, M. Sousa, P. Vieira, M. P. Queluz, and A. Rodrigues, "A no-reference user centric qoe model for voice and web browsing based on 3g/4g radio measurements," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2018.
[18] A. R. Reibman, V. A. Vaishampayan, and Y. Sermadevi, "Quality monitoring of video over a packet network," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 327–334, 2004.
[19] W. Robitza, A. Ahmad, P. A. Kara, L. Atzori, M. G. Martini, A. Raake, and L. Sun, "Challenges of future multimedia qoe monitoring for internet service providers," *Multimedia Tools and Applications*, vol. 76, no. 21, pp. 22243–22266, 2017.
[20] ITU, "ITU-T Recommendations P.1203, Avail.: goo.gl/xqtvnm," 2017.
[21] 3GPP, "LTE: Physical layer measurements. TS 36.214, release 10," 2013.
[22] 3GPP, "LTE: Physical layer procedures. TS 36.213, release 14," 2017.
[23] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in lte cellular networks: Key design issues and a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 678–700, 2013.
[24] F. Afroz, R. Subramanian, R. Heidary, K. Sandrasegaran, and S. Ahmed, "Sinr, rsrp, rssi and rsrq measurements in long term evolution networks," *International Journal of Wireless & Mobile Networks*, 2015.
[25] R. Zhohov, D. Minovski, P. Johansson, and K. Andersson, "Real-time performance evaluation of LTE for IIoT," in *The 43rd IEEE Conference on Local Computer Networks (LCN)*, IEEE, 2018.
[26] K. Hedayat, R. Krzanowski, A. Morton, K. Yum, and J. Babiarz, "A two-way active measurement protocol (twamp)," tech. rep., 2008.
[27] InfoVista, "Tems Pocket, Available: goo.gl/pjhngl," Dec. 2018.
[28] PIMM-DMA, "Available: sics.se/projects/pimm," Dec. 2018.
[29] FFMPEG, "Library, available: ffmpeg.org," Dec. 2018.
[30] C. Bernardin, "Methods and systems for selecting drive routes for testing RF coverage in a radiotelephone system," 2000. US Patent 6,052,583.
[31] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging artificial intelligence applications in computer engineering*, vol. 160, pp. 3–24, 2007.
[32] ITIL, "SLA: Glossary of terms, definitions and acronyms, v3. available: goo.gl/kzdb7p," Dec. 2018.
[33] A. Khan, L. Sun, E. Jammeh, and E. Ifeachor, "Quality of Experience-driven adaptation scheme for video applications over wireless networks," *IET communications*, vol. 4, no. 11, pp. 1337–1347, 2010.
[34] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," *The Netflix Tech Blog*, vol. 6, 2016.
[35] F. Pitié, D. Kelly, T. Foucu, N. Harte, and A. Kokaram, "Assessment of audio/video synchronisation in streaming media," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 171–176, IEEE, 2014.
[36] scikit learn, "Random Forest. available: goo.gl/cpwqds," Dec. 2018.
[37] P. E. Utgoff, "Incremental induction of decision trees," *Machine learning*, vol. 4, no. 2, pp. 161–186, 1989.
[38] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-agnostic interpretability of machine learning," *arXiv preprint arXiv:1606.05386*, 2016.
[39] A. Inc, "Us9185587b2 patent: Load estimation in 3gpp networks," 2015.
[40] J. N. van Rijn, S. M. Abdulrahman, P. Brazdil, and J. Vanschoren, "Fast algorithm selection using learning curves," in *International symposium on intelligent data analysis*, pp. 298–309, Springer, 2015.