2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) | 978-1-7281-7586-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/PIMRC50174.2021.9569490

Network under Control: Multi-Vehicle E2E Measurements for AI-based QoS Prediction

Alexandros Palaios*, Philipp Geuer*, Jochen Fink†, Daniel F. Külzer‡, Fabian Göttsch§, Martin Kasparick†, Daniel Schäufele†, Rodrigo Hernangómez†, Sanket Partani¶, Raja Sattiraju¶, Atul Kumar§, Friedrich Burmeister§, Andreas Weinand¶, Christian Vielhaus‖, Frank H. P. Fitzek‖, Gerhard Fettweis§, Hans D. Schotten¶, Sławomir Stańczak†**

*Ericsson Research, Germany, {alex.palaios, philipp.geuer}@ericsson.com

†Fraunhofer Heinrich Hertz Institute, Germany, {firstname.lastname}@hhi.fraunhofer.de

‡BMW Group Research, New Technologies, Innovations, Germany, daniel.kuelzer@bmwgroup.com

§Vodafone Chair, Technische Universität Dresden, Germany, {firstname.lastname}@tu-dresden.de

¶Technische Universität Kaiserslautern, Germany, {partani, sattiraju, weinand, schotten}@eit.uni-kl.de

‖Deutsche Telekom Chair, Technische Universität Dresden, Germany, {firstname.lastname}@tu-dresden.de

**Network Information Theory Group, Technische Universität Berlin, Germany

***Abstract*—In the future, mobility use cases will depend on precise predictions, with Quality of Service (QoS) prediction being a prominent example. This paper presents realistic measurements from today's vehicles to support robust QoS prediction in the future. Based on a dedicated and controlled measurement campaign, we highlight aspects of the wireless environment and the device characteristics, like the sampling rates, that influence the collected datasets. If not properly handled, such characteristics might hinder the performance of Artificial Intelligence-based algorithms for QoS prediction. Therefore, we also provide insights on dataset characteristics that should be further used to enable easier adoption of AI-based algorithms. New AI-based algorithms should be able to operate in very diverse radio environments with data captured from different devices. We provide several examples that highlight the importance of thoroughly understanding the datasets and their dynamics.**

***Index Terms*—Artificial Intelligence, Machine Learning, Quality of Service Prediction, E2E Measurements, Network Dynamics**

I. Introduction

Artificial Intelligence (AI) is expected to greatly enhance the efficiency, flexibility, and proactivity of modern communication systems [1]. By leveraging Machine Learning (ML) methods, AI can learn from experience and solve problems without being explicitly programmed. Replacing conventional wireless design concepts with ML mechanisms is a promising means to handle the ever-increasing complexity of today's networks [2]. ML techniques are enablers for the automation of network functions, ensuring efficient management of network resources and flexibility to meet user demands [3].

In recent years, a variety of new use cases have emerged in the field of vehicular communications, including connected autonomous driving, platooning, cooperative maneuvering, teleoperated driving, and smart navigation [4], [5]. Each of these use cases has individual Quality of Service (QoS) requirements, including data rate, latency, or reliability constraints [4], which must be satisfied to ensure safe operation. In wireless applications with high mobility, QoS metrics can change drastically within short time periods [6]. This calls for a QoS prediction, which allows the application to adapt proactively to such changes.

Because QoS parameters are influenced by a vast diversity of factors, their prediction usually relies on ML methods. However, these methods are highly dependent on the training datasets, and imbalanced or insufficient data might severely degrade the QoS prediction's performance. Therefore, data generation, selection, and pre-processing need to be designed very carefully.

Owing to the complexity of wireless systems, measurement campaigns do typically not capture all relevant features influencing certain QoS parameters. This work aims at overcoming some limitations of previous studies by conducting a measurement campaign with a broad scope in a highly controllable cellular network. We collected data from the network and User Equipment (UE) under different measurement scenarios with high temporal resolution. Different traffic types and patterns as well as diverse levels of background traffic are employed, for both vehicle-to-network-to-vehicle and vehicle-to-network communication.

The remainder of this paper is organized as follows. Section II summarizes the state of the art in data generation for QoS prediction. Section III describes the measurement campaign. In Section IV we give an overview of selected results, looking at data characteristics that are important for ML algorithms. Finally, we draw conclusions in Section V.

II. State of the Art

Measurement campaigns for QoS prediction described in the literature are often based on a single device type that collects measurements over different areas and time periods [7]. However, the captured radio environment characteristics can vary significantly with varying hardware characteristics, e.g., antenna types and placements. To which extent this influences

This work was supported by the Federal Ministry of Education and Research (BMBF) of the Federal Republic of Germany as part of the AI4Mobile project (16KIS1170K). The authors alone are responsible for the content of the paper.

This document is a preprint of: A. Palaios, P. Geuer, J. Fink et al., "Network under Control: Multi-Vehicle E2E Measurements for AI-based QoS Prediction," in Proceedings of IEEE International Symposium On Personal, Indoor And Mobile Radio Communications (PIMRC 2021), Virtual Conference, (pp. 1432-1438), Sep 2021. DOI:10.1109/PIMRC50174.2021.9569490

the performance of ML models is still unknown, particularly when such models should be used on different device types. In many cases, the data sampling interval of the user equipment is in the order of seconds (e.g., [8], [9]). Such low sampling rates may be insufficient for use-cases with high mobility [10], [6] where the terminal can experience drastically different radio dynamics (e.g. while entering a tunnel).

Several measurement campaigns [11], [12] rely on stationary measurement positions that do not capture precisely the radio environment over different locations. It is known that the radio environment can change drastically within a few meters [13]. Therefore, vehicular speed constitutes a critical parameter for the accuracy of the QoS prediction. Several measurement campaigns have been conducted with vehicles traveling up to 100 km/h (e.g., [6], [14]). While some authors have investigated the impact of mobility on QoS metrics such as throughput (see, e.g., [8]), the influence of vehicular velocity and the sampling frequency on the feature distributions is still largely unknown.

Most studies rely on public networks. Such datasets incorporate unknown network dynamics that cannot be resolved easily. This leads to an additional uncertainty factor that can limit the effectiveness of ML algorithms [15]. Data collection typically happens only at the UE side, resulting in a lack of network-based information. There are a few exceptions where authors have tried to estimate the network status, with an example being the estimation of the base station utilization [10], [11]. It is still an open question to what extent network-based information can improve the performance of ML prediction algorithms [16]. Another limitation is that measurement campaigns are typically conducted during similar time periods (i.e. working hours), with multiple features being relatively stable, such as interference that changes slowly over time [17].

While some studies (e.g. [11], [18]) use different traffic patterns and protocols (Transmission Control Protocol (TCP), User Datagram Protocol (UDP)) to account for different use cases, the effects of Vehicle-to-Network (V2N) and Vehicle-to-Network-to-Vehicle (V2N2V) communication have not yet been studied together in one measurement campaign. Measurements that capture more dynamic and complex environments can provide a more in-depth understanding of the applicability of ML and its capability to generalize on different radio environments, compared to smaller-scale testbeds, as in [6], [14]. As there is a relative lack of publicly available data, studies considering QoS prediction with multiple UEs, e.g., [19], [20], typically resort to simulations. However, simulated data suffers from imperfections (limited dynamics and stationary distributions) that can lead to underperforming ML models. To acquire real-world training data for predictions leveraging measurements from other vehicles, e.g., using information from a preceding vehicle to improve QoS prediction, measurement campaigns with multiple vehicles are required.

With the measurement campaign described in Section III, we aim to provide new insights on several aspects mentioned above. We consider different mobility scenarios in highway, urban and rural environments with different measurement devices.

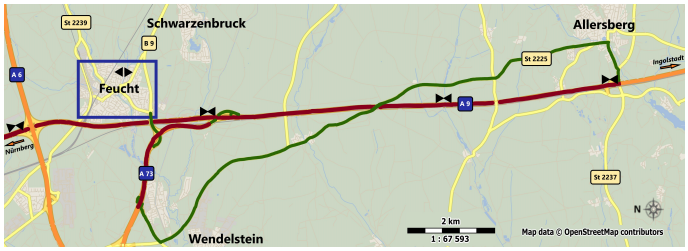

Fig. 1. The map of the areas where measurements took place. The segments of the highway are highlighted with dark red, the rural and side street areas with dark green. The blue frame highlights the suburb Feucht. The black triangles are the base station cells, showing also the coverage direction.

The sampling intervals of the devices range from milliseconds to seconds. Different traffic types and patterns as well as diverse levels of background traffic are employed, for both V2N and V2N2V communication. In addition to the mobile measurement devices, we can access network information from the core and base stations, and we are able to control interference caused by cell load.

III. Measurement Campaign

Within this work, we aim to overcome the described limitations in literature and capture information from all network nodes at a fine granularity. The measurement campaign was performed in the German 5G-ConnectedMobility test field[1] over one week with four vehicles to span V2N and V2N2V communication. In total, a distance of around 3300 kilometers was covered. The network control allowed to investigate arbitrary traffic patterns, regulate interference, and, e.g., reconstruct live base station data. In the following, we describe the measurement area, the network and measurement setup, and introduce the considered measurement scenarios.

A. Measurement Area

The 5G-ConnectedMobility test field is located in the south of Nuremberg, Germany. Figure 1 gives an overview of the measurement area. The propagation scenarios comprised high-speed highway driving, suburban commuting, and rural roads. The highway measurements between Allersberg and the highway junction Nuremberg East, containing a distance of approximately 18 km of continuous highway section (highlighted in red in Fig. 1) experienced special focus during the measurement campaign. Moreover, two base stations covered the suburb Feucht (framed in blue in Fig. 1), with an area of about 10 km^2 and approximately 14,000 inhabitants. It is located close to the highway and allows for suburban scenario measurements. Next to the highway, rural roads connect the town of Allersberg with Feucht, a diverse propagation environment with forests and small towns in the countryside (marked in green in Fig. 1).

B. Measurement Setup

The mobile radio network is a private network covering the test area with five base stations operating at 700 MHz. As seen in Figure 1, four base stations are positioned along the highway,

[1] http://www.5g-connectedmobility.com/

This document is a preprint of: A. Palaios, P. Geuer, J. Fink et al., "Network under Control: Multi-Vehicle E2E Measurements for AI-based QoS Prediction," in Proceedings of IEEE International Symposium On Personal, Indoor And Mobile Radio Communications (PIMRC 2021), Virtual Conference, (pp. 1432-1438), Sep 2021. DOI:10.1109/PIMRC50174.2021.9569490

and one base station is located in the suburban city of Feucht. Each base station covers two cells resulting in good coverage at a part of the A9 highway, Feucht, and the rural areas close to the highway. The network employs Long-Term Evolution (LTE)-Frequency Division Duplexing (FDD) with a respective bandwidth of 10 MHz. The base stations can also be configured to enable artificial cell load. If enabled, a configured fraction of downlink Physical Resource Blocks (PRBs) is loaded with dummy data, resulting in increased interference, especially in neighboring cells. Figure 2 depicts the network architecture and our measurement methodology. The base stations are connected to a core network, which is split between different locations. The central cloud, including functionalities as Mobility Management Entity (MME) or Home Subscriber Server (HSS), is located in Aachen, while Mobile Edge Computing (MEC) provides the Packet Data Network Gateway (PGW) and a local server close to the base stations. All parts of the radio and core network are synchronized via GPS.

All cars were equipped with an identical Dedicated Measurement Equipment (DME) for fine-grained mobile network analytics. Moreover, two cars carried Commercial Off-The-Shelf (COTS) mobile phones for comparison measurements. Additionally, six COTS devices distributed over the cars solely acted as background traffic data generators to populate an otherwise empty network.

An overview of the captured data from all network nodes, including the devices, the base stations, the core, and additional sources, is given in Table I. Each DME consists of a small form factor PC with an Intel Core i7-7500U CPU and a Linux-based operating system. Moreover, it includes a (Sierra Wireless AirPrime MC7430) LTE modem connected to a roof-mounted 2x2 Multiple-Input Multiple-Output (MIMO) car antenna. Additionally, a roof-mounted GPS receiver (Garmin GPS 18x LVC) connected to the DME provided accurate location and time synchronization via Pulse Per Second (PPS)-signal. We used the application Iperf (acting as a client) on all devices inside the cars to generate data in Downlink (DL) or Uplink (UL) direction, while several server instances are running on the local server provided by MEC. Each device and the local server logged the requested and achieved throughput, respectively. The COTS measurement devices use the Android API to capture Physical Layer measurements at a sampling rate of 1 Hz, including, e.g., Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), and the connected cell, as well as the location. On the DME, the open-source software MobileInsight [21] is used to capture all modem-chipset-specific messages from the full LTE stack with sampling intervals of up to 1 ms (see Table I). All individual incoming and outgoing packets are collected from the network interface associated with the LTE modem. The data generators only log the connected base station in addition to the Iperf logs.

Furthermore, data was also collected at the base stations and the core network. The base stations generated the typical 15-minute counters. Due to our holistic network control, we can use our information about the connected cells and the throughput at the device side to compute, e.g., cell load and other reconstructed real-time information for considerably shorter sampling intervals of approximately 50 ms. Additionally, we capture MME traces from the core network, which, for example, indicate handover procedures and paging. At the PGW, we record all individual packets from all devices with time stamps, enabling us to compute the one-way latency between DMEs and the local server, among other metrics such as throughput or consecutive packet losses at fine time granularity.

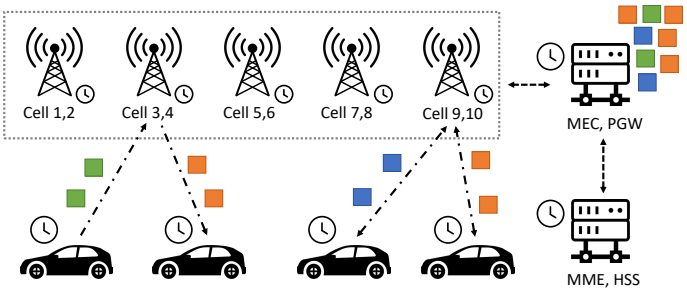

Fig. 2. A simplified view of the measurement methodology used. Vehicles and server generate marked traffic (green: low, blue: medium, orange: maximum throughput). At the packet gateway the marked packets enable new features, like current load at base stations with fine time granularity. All clocks in all network nodes are synchronized.

C. Measurement Scenarios

The measurement campaign focused on collecting datasets in a way such that effects of specific measurement parameters can be studied in-depth. In Figure 3, we show the parameters studied. At a high level, measurement parameters are grouped in five categories that include the configuration of base stations, the data generation, the type of network connectivity, the scenarios covered, and the mobility schemes. Three mobility schemes were investigated: 1. leading car including a sequence of cars driving in the same route with a fixed time lag to the leading car, 2. moving and stationary schemes that included two vehicles in the same cell where one vehicle is parked and the second vehicle moving in and around the same cell, 3. stationary measurements where all cars were parked. The data generation parameters included different patterns where devices are competing to achieve maximum throughput from the radio interface and also with a mixture of traffic types. In total we executed 31 different measurement scenarios, with each scenario having a specific focus (i.e. studying the effects of one of the parameters by keeping all other parameters fixed). A typical measurement scenario lasted between 40 to 60 minutes, with the stationary measurements lasting typically a few hours. To further reduce downtime between different measurement scenarios, special software was developed enabling fast configuration of the devices between the measurement runs.

IV. Measurement Evaluation

This section covers our first results of this measurement campaign. We have specifically focused on presenting and discussing characteristics that can improve or hinder adoption of ML algorithms for QoS prediction.

This document is a preprint of: A. Palaios, P. Geuer, J. Fink et al., "Network under Control: Multi-Vehicle E2E Measurements for AI-based QoS Prediction," in Proceedings of IEEE International Symposium On Personal, Indoor And Mobile Radio Communications (PIMRC 2021), Virtual Conference, (pp. 1432-1438), Sep 2021. DOI:10.1109/PIMRC50174.2021.9569490

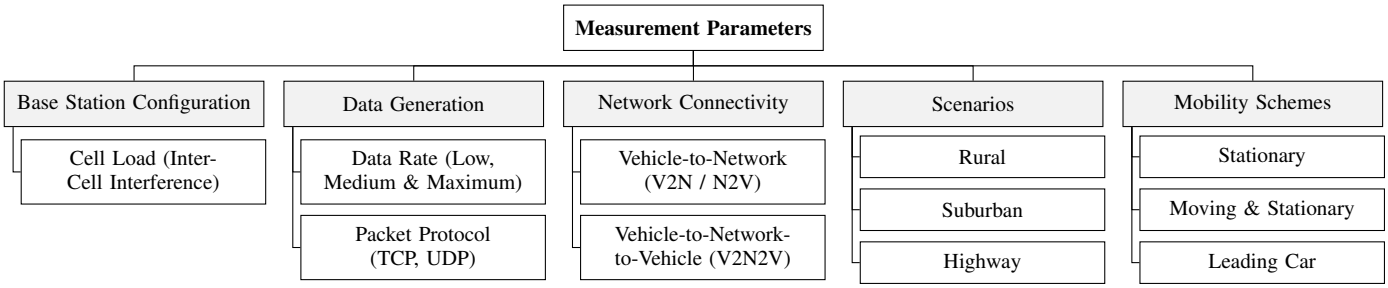

Fig. 3. The configurable parameters for the measurement campaign. Different combinations of the leaf nodes would result in different measurement configurations per device. We selected a subset of these combinations so as to cover the entire spectrum of available configuration parameters.

TABLE I
OVERVIEW OF CAPTURED DATA FROM ALL NETWORK NODES.

Network Node	Information Type	Sampling Interval	Features (Examples)	QoS Metrics (Examples)
DME	Physical Layer (PHY)	10 ms	RSRP, RSRQ, CQI	
	Medium Access Control (MAC)	40 ms	Buffer Status Report (BSR)	
	Packet Data Convergence Protocol (PDCP)	0.5 s	PDCP UL & DL Statistics	
	Radio Link Control (RLC)	45 ms	RLC transmissions & receptions	
	Radio Resource Control (RRC)	Event-based	OTA Packet Log	
	Location (via GPS)	1 s	Longitude, Latitude, Speed	
	Packet Capturing Tool (using tcpdump)	1 ms		DL Throughput & Latency
COTS	Physical Layer (PHY)	1 s	RSRP, RSRQ	
	Measurement Tool (using Iperf)	1 s		DL Throughput
Base Station	Standard Operator Averages	15 min	Interference, Cell Load	
	Reconstructed Real-time Information	50 ms	Interference, Cell Load	
Core	Mobility Management Entity (MME) Traces	Event-based	Handover	
	Packet Capturing Tool (using tcpdump)	1 ms		UL Throughput & Latency
Database	Weather & Traffic APIs	Up to 1 min	Traffic Flow, Precipitation	

A. *Radio Environment*

As a first analysis, we take a look at the radio environment characteristics in the different measurements areas. In Figure 4, we provide the Probability Density Functions (PDFs) of the received RSRP values of the DME in vehicle 3, aggregated over all measurement scenarios, and separated based on three different radio environments: the highway, the suburban area (Feucht), and the rural areas around the highway. The network is deployed with a focus on providing great coverage to the highway, that is visible by the higher RSRP values in Figure 4. The RSRP values within the extensive coverage area enable the support of very demanding use cases. The highway is a relatively stable radio environment with line-of-sight (LOS) conditions for the most of it, producing a more uniform type of distribution. On the other hand, in the suburban city of Feucht, there are richer radio dynamics that can be explained by the continuous transition of LOS and non-line-of-sight (NLOS) conditions. These are typically found in suburban environments as vehicles travel between the building blocks. In the rural area, the presence of trees partially block the LOS producing a much smoother RSRP distribution compared to the one from the suburban area. It is noticeable that the direction of the antennas was optimized for highway coverage and not for the rural areas. Nevertheless, we experienced good coverage.

Another interesting remark from Figure 4 is that, as described in [22], the reporting range of an LTE receiver is defined from -156 dBm to -44 dBm and that we could capture a big part

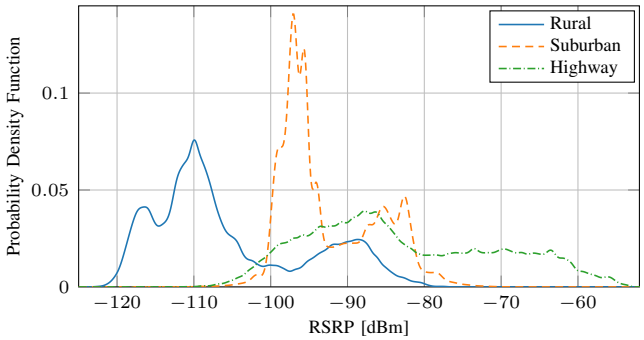

Fig. 4. The PDFs of the received RSRP in three different radio environments, where the RSRP captured in all measurement scenarios by the DME in vehicle 3 is considered.

of that range of values since we covered very diverse radio environments. If a measurement campaign were carried out in a fixed radio environment, it might only capture a small set of different values. This limited range of input could cause issues to the ML algorithm, which might not generalize well outside of the range of the training input. Another point is that if the measurement campaign does not cover the whole input range of the features, there is a risk that non-linearities between the input features are not successfully captured. We will come back to this point in the next subsections.

B. *Impact of Different Device Types*

Since studies are typically conducted with similar types of devices, there is a need to understand to which extent devices

This document is a preprint of: A. Palaios, P. Geuer, J. Fink et al., "Network under Control: Multi-Vehicle E2E Measurements for AI-based QoS Prediction," in Proceedings of IEEE International Symposium On Personal, Indoor And Mobile Radio Communications (PIMRC 2021), Virtual Conference, (pp. 1432-1438), Sep 2021. DOI:10.1109/PIMRC50174.2021.9569490

influence the characteristics of collected datasets. Diverse devices have different radio transceivers, antenna positions and sampling speeds. In Figure 5, we focus on two devices in the same car that are measuring the RSRQ. The x axis here shows the vehicle's latitude. Due to the north-south orientation of the A9 highway, the latitude coordinates $49.25\,°\,\mathrm{N}$ to $49.40\,°\,\mathrm{N}$ roughly map into $397.5\,\mathrm{km}$ to $380.7\,\mathrm{km}$ on said highway. The two devices are a DME and a COTS that are inside the car. As expected, the COTS is not able to capture as high values as the DME. The DME has an external antenna and thus receives higher values, where RSRQ values larger than $-5\,\mathrm{dB}$ typically refer to excellent quality. Moreover, it is interesting that the COTS does not capture the severe short time-scale deterioration of the channel quality, even though it is inside the car. RSRQ values below $-12\,\mathrm{dB}$ indicate unusable signal quality. Such limited captured dynamics might be a compound effect of lower sampling speeds of the COTS and possibly, averaging time windows that are typically employed in such devices. This shows to some extent the importance of accounting for specific device limitations, specifically for ML applications that require higher accuracy. On the other hand, the COTS is capturing quite precisely the average evolution of the RSRQ values. The addition of an appropriate device error margin could compensate for the loss of resolution that lower-cost consumer devices typically suffer from.

Figure 6 gives another example of the impact of different device types on the perceived radio environment. In this case, we consider the Cumulative Density Functions (CDFs) of measured RSRP values and compare identical device types in two different vehicles for a leading car highway measurement. We observe that COTSs have integer resolution (resulting in a step function), while DMEs record Physical (PHY) layer parameters at a higher resolution. Both vehicles capture similar radio environments, as their CDF shapes are similar. However, the DME curves are shifted due to factors such as different antenna positions on different cars. In vehicle 4, the COTS shows a comparable dynamic to that of the DMEs, while vehicle 3 COTS measures RSRP values with a higher probability of poor signal quality compared to vehicle 3 DME. The latter COTS was positioned on the car's rear bench seat compared to a window fixation of vehicle 4 COTS. Therefore, the position of the device antennas influences the captured RSRP measurements. Finally, Figure 6 confirms the earlier insight of highway propagation scenarios being advantageous for ML because of a single, steep slope around $-95\,\mathrm{dBm}$ for vehicle 3 DME and $-105\,\mathrm{dBm}$ vehicle 4 DME, respectively. In any case, we observe that different devices capture similar dynamics, at least to some extent. This indeed hints at the reuse of prior measurement data by other devices.

C. Temporal Analysis of Fading Effects

The design of the measurement campaign allows to augment the QoS prediction with features measured by other vehicles. To investigate the feasibility of this approach, we consider the time series of PHY layer features measured by devices in two vehicles travelling $3.05\,\mathrm{min}$ apart on average. Figure 7

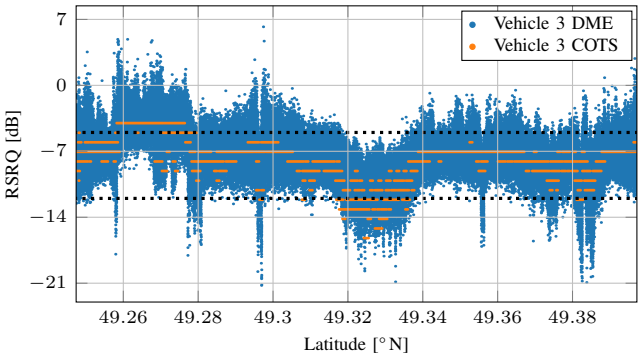

Fig. 5. The RSRQ as captured from two different device types inside the same vehicle. Dashed lines at $-5\,\mathrm{dB}$ and $-12\,\mathrm{dB}$ represent a typical operating range.

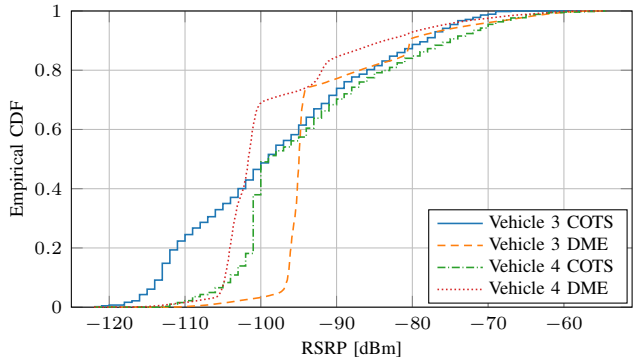

Fig. 6. Example for the impact of the devices on the measurement data.

shows the normalized cross-correlation between the Received Signal Strength Indicator (RSSI), RSRP, and RSRQ time series recorded by the two vehicles for each feature during one highway measurement run of $1\,\mathrm{h}$ duration. As all time series are resampled to intervals of $1\,\mathrm{s}$, this analysis mainly captures the effects of large-scale fading. It can be seen that all curves reach their maximum at a temporal lag that corresponds to the average lag between the vehicles' trajectories. Similar effects can be observed for all measurement runs with cars following the same route. This indicates that measurements from preceding vehicles can be valuable features for predicting the propagation environment. The incorporation of such features and their potential benefits to QoS prediction will be studied in future publications.

While the large-scale fading shows to be correlated across vehicles, we would like to get insights into the stationarity of the small-scale fading. Figure 8 shows the median Kullback-Leibler (KL) divergence between time-delayed vehicles on the same route as in Figure 7, where we estimate mean-adjusted RSRP distributions over intervals of $5\,\mathrm{s}$ using the method described in [23]. This interval length is motivated by average stationarity intervals in highway scenarios [13], while guaranteeing sufficient data per density estimation. The choice of the median stems from the fact that it is unaffected by nonstationarities (if less than half of the distributions are affected). We observe distinct minima at the average time lags between the vehicles, with the effect being less pronounced for

This document is a preprint of: A. Palaios, P. Geuer, J. Fink et al., "Network under Control: Multi-Vehicle E2E Measurements for AI-based QoS Prediction," in Proceedings of IEEE International Symposium On Personal, Indoor And Mobile Radio Communications (PIMRC 2021), Virtual Conference, (pp. 1432-1438), Sep 2021. DOI:10.1109/PIMRC50174.2021.9569490

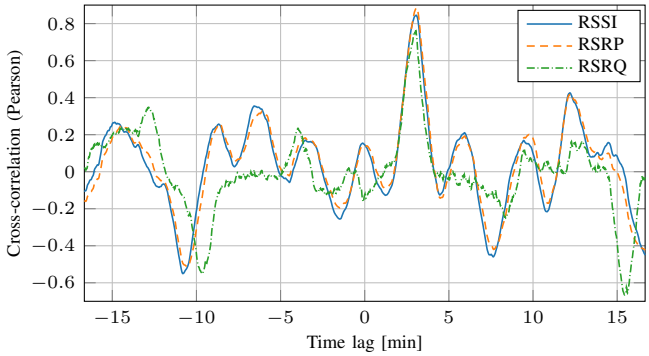

Fig. 7. Normalized cross-correlation of the RSSI, RSRP, and RSRQ captured by two vehicles travelling 3.05 min apart from each other on average.

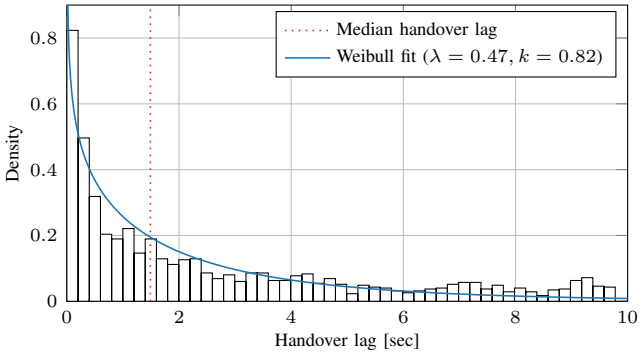

Fig. 9. Histogram of highway inter-base station handover lags for handovers within 10 s of the first handover in a vehicle (using the core's MME traces).

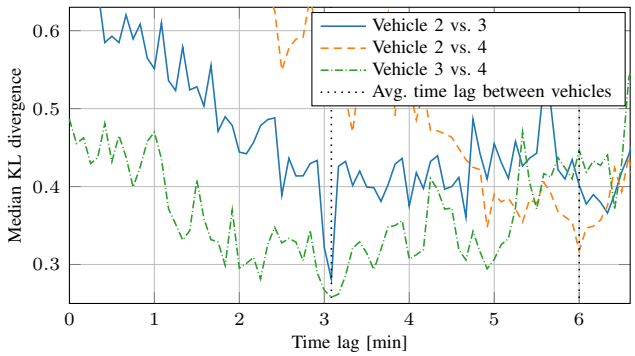

Fig. 8. KL divergence for different time lags between the zero-mean RSRP PDFs of two vehicles driving the same route delayed in time.

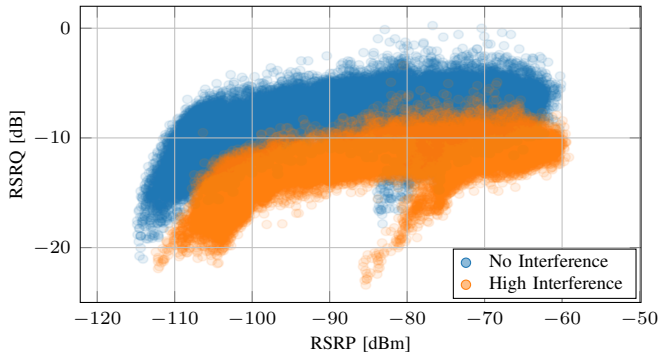

Fig. 10. An example on the collected PHY layer input features drift. The increased cell load increases the experienced interference at the terminals.

further-delayed vehicles. These results hint at fast-fading distributions remaining stationary or at least relatively unchanged over several minutes in spatial regions.

D. Handover Characteristics

Within the context of Vehicle-to-Everything (V2X) communication, many use cases require a very low latency in the order of 50 ms. We noticed that devices inside the same vehicle perform handovers at different times, often delayed by considerably more than 50 ms. To study this effect in more depth, we consider the device behavior per individual vehicle (including DME, COTS, and traffic generators) and leverage the core's MME traces for millisecond resolution on inter-E-UTRAN Node B (eNB) handovers perceptible by the core. We define the handover lag as the time interval between the first device switching to a new cell in the respective car and any of the remaining devices' handover occurrence. Specifically, all time lags within a window of 10 s are counted. Figure 9 shows the handover lag histogram aggregated over all highway measurement runs and vehicles, which has a mean value of 2.65 s. Moreover, 28 % of handover events were excluded from this statistic, as they exhibited ping-pong effects or occurred isolated from the other devices in the same vehicle and considered time window. In summary, for V2X applications, handover procedures need to be considered towards robust QoS prediction. If the same analysis were to be performed solely based on end-device recordings, the COTS sampling interval of 1 s would introduce a Mean Absolute Error (MAE) of 1.15 s per handover lag.

E. Impact of Cell Load

Studies on ML-based QoS prediction often assume stationary input feature distributions [24], as this simplifies significantly the application of ML algorithms. However, several parameters of wireless networks are known to exhibit non-stationary distributions [17]. For example, the cells are typically highly loaded during peak-hours. The authors of [25] have shown that the dependency between RSRQ and RSRP is not linear and that their correlations change based on the received power. As high cell load results in high interference from other cells, and thus in an increased received power level, we expect the dependency of RSRQ to RSRP to change over time in a live network.

Our measurement setup enables us to emulate cell load in neighboring cells, which serves us to generate interference in an active way. In Figure 10, we show that the relationship between RSRQ and RSRP also changes as a function of the cell load. In fact, we conducted two different measurement runs where we kept all parameters stable and only adjusted the cell load. As a result, it can be seen that the increased cell load shifts the collected points of the two parameters considerably. This can serve as a good example for the high temporal variation of many of the interrelations of the input features collected from radio environments. Still, more in-depth studies are needed to characterize those in more detail. Other than that, we have been able to confirm the non-linear behaviour between those

two inputs features, which implies that a wide range of these measurements must be considered in order to fully capture their relationship. An ML-based algorithm should have access to input features that can successfully track such characteristics to provide robust predictions over long periods of time.

V. Conclusion

In this paper, we presented first insights of a multi-vehicle End-to-End (E2E) measurement campaign showing that more control of the measurement setup can reveal important characteristics of the collected features. Such information will help to shape the communication system of future transport systems with vehicles. We showcased that dedicated and controlled measurement campaigns are needed to capture the complex inter-dependencies and characteristics of the radio environment.

We found that device type, placement, and sampling interval impact measured features, which must be considered when applying ML approaches. Using the same devices and configuration for both training and testing, which is a common practice, may lead to overly optimistic results. By using multiple vehicles, each of them equipped with multiple devices, we could identify several challenges and opportunities for QoS prediction mechanisms. On the one hand, even when placed in the same vehicle, devices tend to show different behaviors during the handover procedure, and this uncertainty must be integrated into E2E prediction mechanisms. On the other hand, our findings indicate that data measured by preceding vehicles may be useful to improve QoS prediction. We plan to investigate further how this information can be incorporated. By capturing information from all parts of the network, we are able to identify more precisely the complex interrelations among features. This can improve the accuracy of prediction mechanisms, as well as their ability to generalize.

In the future, we plan to conduct more measurement campaigns and provide insights on dataset characteristics for higher frequency bands. Moreover, we plan to continue this in-depth analysis, also integrating the QoS measurements and provide new ML algorithms that can handle such characteristics of the dataset.

Acknowledgment

We thank all AI4Mobile partners for their valuable contributions in the planning and execution of the measurement campaign. We would also like to thank Business Area Networks, Business Area Digital Services, and Customer Unit Western Europe from Ericsson. Specifically we thank R. Wellens, L. Garcia, and C. Sous for their active support before, during, and after the measurements.

References

[1] R. Li, Z. Zhao *et al.*, "Intelligent 5G: When cellular networks meet artificial intelligence," *IEEE Wirel. Commun.*, vol. 24, no. 5, pp. 175–183, Oct. 2017.

[2] Y. Huang, C. Xu *et al.*, "An overview of intelligent wireless communications using deep reinforcement learning," *Journal of Communications and Information Networks*, vol. 4, no. 2, pp. 15–29, Jun. 2019.

[3] H. Wu, X. Li, and Y. Deng, "Deep learning-driven wireless communication for edge-cloud computing: opportunities and challenges," *J. Cloud Comput.*, vol. 9, pp. 1–14, Apr. 2020.

[4] D. F. Külzer, M. Kasparick *et al.*, "AI4Mobile: Use cases and challenges of AI-based QoS prediction for high-mobility scenarios," in *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, Apr. 2021.

[5] L. Figueiredo, I. Jesus *et al.*, "Towards the development of intelligent transportation systems," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Aug. 2001, pp. 1206–1211.

[6] L. Torres-Figueroa, H. F. Schepker, and J. Jiru, "QoS evaluation and prediction for C-V2X communication in commercially-deployed LTE and mobile edge networks," in *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, May 2020, pp. 1–7.

[7] C. Yue, R. Jin *et al.*, "Linkforecast: cellular link bandwidth prediction in LTE networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 7, pp. 1582–1594, Jul. 2018.

[8] A. Narayanan, E. Ramadan *et al.*, "Lumos5G: Mapping and predicting commercial mmWave 5G throughput," in *Proc. ACM Internet Measurement Conference (IMC)*, Oct. 2020, pp. 176–193.

[9] A. Kulkarni, A. Seetharam, A. Ramesh, and J. D. Herath, "Deepchannel: Wireless channel quality prediction using deep learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 443–456, Jan. 2020.

[10] J. Lee, S. Lee *et al.*, "PERCEIVE: deep learning-based cellular uplink prediction using real-time scheduling patterns," in *Proc. ACM Conf. on Mobile Systems, Appl., and Services (MobiSys)*, Jun. 2020, pp. 377–390.

[11] R. Falkenberg, K. Heimann, and C. Wietfeld, "Discover your competition in LTE: Client-based passive data rate prediction by machine learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–7.

[12] N. Suga, K. Yano *et al.*, "Estimation of probability density function using multi-bandwidth kernel density estimation for throughput," in *Proc. Int. Conf. on Artif. Intell. in Inf. and Communication (ICAIIC)*, Feb. 2020, pp. 171–176.

[13] L. Bernadó, T. Zemen *et al.*, "The (in-)validity of the WSSUS assumption in vehicular radio channels," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2012, pp. 1757–1762.

[14] J. Schmid, M. Schneider, A. HöB, and B. Schuller, "A deep learning approach for location independent throughput prediction," in *Proc. IEEE Int. Conf. on Connected Vehicles and Expo (ICCVE)*, Nov. 2019, pp. 1–5.

[15] D. Schäufele, M. Kasparick *et al.*, "Terminal-side data rate prediction for high-mobility users," in *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, Apr. 2021.

[16] B. Sliwa, R. Falkenberg, and C. Wietfeld, "Towards cooperative data rate prediction for future mobile and vehicular 6G networks," in *Proc. 6G Wireless Summit (6G SUMMIT)*, Mar. 2020, pp. 1–5.

[17] A. Palaios, J. Riihijärvi, O. Holland, and P. Mähönen, "A week in London: Spectrum usage in metropolitan London," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2013, pp. 2522–2527.

[18] Q. Xu, S. Mehrotra, Z. Mao, and J. Li, "PROTEUS: network performance forecast for real-time, interactive mobile applications," in *Proc. ACM Int. Conf. on Mobile Systems, Appl., and Services (MobiSys)*, Jun. 2013, pp. 347–360.

[19] T.-Y. Park, J.-W. Han, and E.-K. Hong, "UE throughput guaranteed small cell on/off algorithm with machine learning," *J. Commun. Netw.*, vol. 22, no. 3, pp. 223–229, Jun. 2020.

[20] D. C. Moreira, I. M. Guerreiro *et al.*, "QoS predictability in V2X communication with machine learning," in *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, May 2020, pp. 1–5.

[21] Y. Li, C. Peng *et al.*, "Mobileinsight: Extracting and analyzing cellular network information on smartphones," in *Proc. ACM Annu. Int. Conf. on Mobile Comput. and Netw. (MobiCom)*, Oct. 2016, pp. 202–215.

[22] *Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements for support of radio resource management*, 3GPP TS 36.133, Rev. 17.0.0, Dec. 2020.

[23] Z. I. Botev, J. F. Grotowski, D. P. Kroese *et al.*, "Kernel density estimation via diffusion," *Ann. Stat.*, vol. 38, no. 5, pp. 2916–2957, Aug. 2010.

[24] K. Uludağ and Ö. Korçak, "Energy and rate modeling of data download over LTE with respect to received signal characteristics," in *Proc. Int. Telecommun. Networks and Appl. Conf. (ITNAC)*, Nov. 2017, pp. 1–6.

[25] V. Raida, M. Lerch, P. Svoboda, and M. Rupp, "Deriving cell load from RSRQ measurements," in *Proc. Network Traffic Measurement and Analysis Conf. (TMA)*, Jun. 2018, pp. 1–6.