V2V- and V2X-Communication data within a distributed computing platform for adaptive radio channel modelling

Florian Pinzel*, Jörg Holfeld◇, Andrej Olunczek◇, Paul Balzer†, Oliver Michler*

* Chair of Transport Systems Information Technology, Technische Universität Dresden, Germany
Email: {florian.pinzel, oliver.michler}@tu-dresden.de

◇ Fraunhofer Institute for Transportation and Infrastructure Systems IVI, Dresden, Germany
Email: {joerg.holfeld, andrej.olunczek}@ivi.fraunhofer.de

† MechLab Engineering UG, Dresden, Germany
Email: balzer@mechlab-engineering.de

***Abstract*—This paper presents findings of the collection and exploitation process of V2X communication data with the aim of developing a measurement data-based radio channel model for the ITS frequency range around 5.9 GHz. Collected under real world conditions, connectivity quality measurements of ETSI ITS G5 communication data form the basis of the presented model prototype. The paper provides insight into the installation and configuration of the communication hardware used. Furthermore, the transmission process of accumulated as well as live data from the vehicles to a big data platform using the IoT message protocol MQTT is investigated. There, the communication data is enriched with other geographically referenced open source data. Finally, the development of a prototype V2X radio channel model using a machine learning process is presented. The model is a helpful instrument for predicting reception qualities in the ITS radio range for previously unknown receiver positions and thus a prerequisite for two exemplary presented use cases.**

***Index Terms*—Vehicle-to-X Communication, ETSI ITS G5, DSRC, Tensorflow Regression Network**

I. INTRODUCTION

A. Problem Formulation

With increasing complexity of Vehicle-to-Anything (V2X) applications the need to ensure robustness and integrity over the entire transmission path increases as well. The use of so-called radio propagation models enables the evaluation of characteristics and limiting factors of a propagation channel. Due to its economic importance, a considerable part of the research work in this area is concerned with cellular mobile radio systems using stationary infrastructure (eNodeB for LTE, BTS for GSM). A significant part of all established radio channel models is therefore only valid for mobile radio scenarios with long distances and lower radio frequencies. However, the propagation characteristics of a V2X radio channel differ significantly from the conditions prevailing in the mobile radio channel. In order to guarantee the performance and reliability of a V2X communication system, special metrics relevant for the V2X radio channel have to be identified, which in turn result in special channel models.

B. Related Work

The topic of radio channel modelling in the frequency range around 5.9 GHz has so far been addressed in different forms. Investigations of *Frankiewicz et al.* aim at the location determination of Intelligent Transportation Systems (ITS) roadside stations for the construction of a large-scale research facility in Brunswick, Germany. [1] Measurement campaigns in real traffic environment, in which both transmitter and receiver are dynamically moved, determine the works of *Cheng et al.* and *Sen et al.* [2], [3] *Cheng's* studies allow statements on the effects of small scale fading to be made with investigations on Doppler spread (spectrum widening due to different Doppler shifts of the multipath paths) and coherence time (time span in which the dispersion has no effect). Instead of Doppler spread, *Sen et al.* consider delay spread (pulse widening due to multipath reception) and frequency correlation by recording the pulse response in the entire broadband channel. *Molisch* provides an excellent summary of the findings of these and other measurement campaigns and publications. [4]

Although all mentioned publications use measurement data from real traffic environment to determine channel properties, the acquired data volume is due to its cost- and time-intensive collection usually limited. In general Measurements are only carried out within a narrow local and temporal framework, impairing informative value and reproducibility of obtained results.

Funded by the federal ministry of transport and digital infrastructure of Germany within the *mFUND* research initiative, https://www.bmvi.de/goto?id=350148

II. FUNDAMENTALS

A. V2X Communication

The core idea of V2X Communication is the creation of so-called Vehicular Ad Hoc Networks (VANETs) between road users or between road users and road side infrastructure for the purpose of data exchange. Such a communication link with low latencies of less than 60 ms and communication ranges of up to 1000 m under optimal conditions enables a multitude of safety and traffic efficiency applications. [5] The ETSI ITS communication standard provides various types of messages for this purpose, such as Cooperative Awareness Message (CAM), Decentralized Environmental Notification Message (DENM), map data (MAP) and Signal Phase and Timing (SPaT). With its ability to also communicate in Non-Line of Sight (NLoS) scenarios, V2X communication is a useful supplement to vehicle sensors with limited spatial effect such as optical, ultrasonic, LIDAR or RADAR sensors.

B. Link budget and radio channel model

A link budget balances the power arriving at the receiver P_{Rx} with the objective of determining the quality of the transmission channel. In a simple form, antenna gains G_{i} are added to the transmission power P_{Tx} in logarithmic form, while connection losses C_{i} and path losses L are subtracted from it:

$$P_{\mathrm{Rx}} = P_{\mathrm{Tx}} + G_{\mathrm{Tx}} - C_{\mathrm{Tx}} - L + G_{\mathrm{Rx}} - C_{\mathrm{Rx}} \tag{1}$$

This form of a link budget can be considered as a starting point for different radio channel models. A radio channel model thus describes the mathematical relationship of the path loss of an electromagnetic signal between transmitter and receiver. For rough estimates of the reception conditions, the modelling of the path loss as free space path loss L_{FSPL} is already sufficient. L_{FSPL} as a ratio between received and transmitted power is subject to distance and frequency dependence.

$$L_{\mathrm{FSPL}} = \frac{P_{\mathrm{Rx}}}{P_{\mathrm{Tx}}} = \left(\frac{c}{4\pi d\ f}\right)^2 \tag{2}$$

From a radio technical point of view, the most important quality parameter within this study is Received Signal Strength Indicator (RSSI). Although the RSSI is in principle a vendor-specific indication, it can be regarded as equivalent to the reception power P_{Rx} at the radio module in this publication. A calibration measurement of the radio modules using a spectrum analyzer did not determine any substantial deviation of the actually received power from the output power of the radio module. This justifies the previous assumption to equate RSSI and P_{Rx}.

C. MQTT Message Protocol

Message Queuing Telemetry Transport (MQTT) is a light weight messaging format based on a publish-subscribe pattern. It has a small footprint regarding network bandwidth and works on top of the standard TCP/IP protocol. Messages will be transferred via MQTT broker. The messages are submitted to a predefined topic while on the other side clients can subscribe to specific topics at these brokers to receive the messages. Developed primarily for Internet of Things (IoT) networks, one of the most significant features of the MQTT protocol is the ability to easily handle disconnections and reconnections due to limited or interrupted network connections. Thus it is also very well suited for the transmission of measurement data from moving vehicles. Broker connection interruptions, e.g. caused by missing mobile connections in tunnels, only result in delayed data transmission.

III. SETUP

A. Data Collection Approach and Vehicle Hardware

Collection of V2X Communication data within the research project is carried out through various measuring platforms. The vast majority of data is collected by two vehicles of the traffic accident research unit of TU Dresden always driving in series. Due to varying operating times, approximately six hours of active driving time per day and the fact that trips are made to urban, suburban and rural areas, the data collected covers a wide range of different propagation conditions. The accident research vehicles as a data source are almost exclusively accountable for V2V communication data. In addition to the vehicles in continuous use, two other vehicles were equipped with compatible communication hardware. These vehicles are particularly suitable for targeted driving and evaluation of special scenarios or areas, such as locations with ITS Road Side Stations. For this reason, an ITS Road Site Station was permanently installed on a building roof within the research project. Figure 1 shows an example of a dedicated measurement run around the specially installed roof antenna. The reduction of the communication connection caused by increased attenuation due to increasing distance and shading of the building becomes clearly visible.

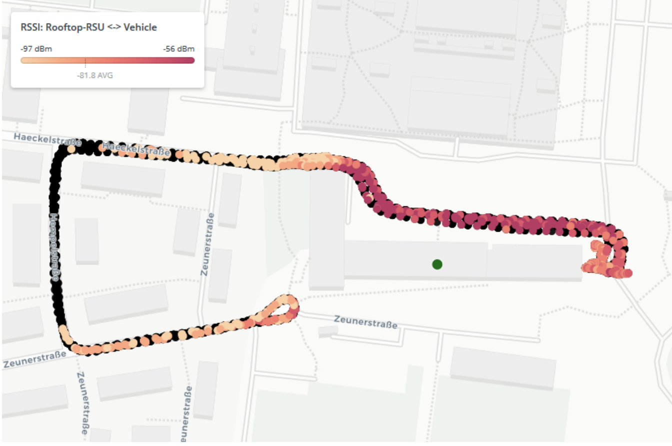


Fig. 1. Decrease in reception power (colored) and communication interruption between RSU (green) and moving vehicle (black)

All vehicles are equipped with three essential components: a computing unit, a radio communication module as well as a integral antenna. The integrated vehicle antennas enable the reception of ITS G5, cellular radio with 2G / 3G / 4G, GNSS technologies and are connected to the V2X communication module. This communication module is connected via Ethernet to an automotive grade computing unit and transmits received ITS G5 messages to it. The used communication module *Cohda Wireless MK5 (Rev. 2.00)* comes with an Global Navigation Satellite System (GNSS) chipset with integrated dead reckoning functionality (*ublox NEO-M8U*). This circumstance promises a robust position accuracy even if the measuring vehicles are located in areas without GNSS reception, e.g. tunnels or urban canyons.

B. Raw Measurement Data

The recorded raw data can be divided into the domains of radio parameters and geographical referencing parameters. Exemplary measured values are shown in Table 1. A unique time stamp serves as an identifier for individual log lines.

TABLE I
Exemplary Raw Data

Value	Example value	Additional note
TIME	1548074554	unix timestamp
ID	1377	collector ID
RSSI	-53 dBm	[23dBm; -97dBm]
TYPE	CAM	ETSI ITS G5 message type
LAT_{Tx}	51.04648	Latitude Ego Vehicle
LON_{Tx}	13.707152	Longitude Ego Vehicle
LAT_{Rx}	51.046433	Latitude Receiver Vehicle
LON_{Rx}	13.706518	Longitude Receiver Vehicle

C. Backend Infrastructure

Data send by the vehicles arrives in telemetry servers that bundle all data messages of all connected vehicle and forward the data messages to a data warehouse and big data platform. Telemetry servers utilize the MQTT protocol (see II-C) for receiving and forwarding data messages. These servers can also be used as edge cloud servers when placed near the infrastructure serving the geographically nearest vehicles.

The big data platform consists of two server clusters, each containing at least three servers. The general structure follows the *lambda architecture*. [6] The first server cluster is responsible for data ingestion, the speed layer and storing filter data. The second cluster is responsible for storing raw data, batch layer and the serving layer. The general structure is depicted in Fig. 2.

A MQTT broker is installed on each cluster server for the purpose of data ingestion. So each telemetry server can choose one of the cluster servers for submitting data messages, depending on availability of the servers. This increases the overall system availability as the loss of one cluster server does not stop data processing. On the speed level cluster *Apache Kafka* is used to manage data streams. A *Kafka Connect* module is responsible for transforming incoming MQTT messages into data streams. Depending on the MQTT topic, the messages are forwarded to two streams, one for batch processing and one for stream processing. The stream processing is implemented as *Kafka Streams* applications and main goal is to evaluate the positions of the vehicles. The applications determine the status of the vehicles (moving/parking) and perform a Hidden-Markov-based map matching to estimate the road where the vehicles are driving. This results in the OpenStreetMap (OSM)-id of the current road section. The data derived from the streaming applications is stored in an *Apache Cassandra* database and is reused to filter raw data. The cluster for the batch layer is a classic *Apache Hadoop* setup. It stores raw data and gives access for analytic evaluation of data. The radio propagation models are developed and tested within this unit. Within the research project it has proved to be advantageous to encapsulate developed propagation models with the help of the container environment *Docker* and to equip it with a dedicated Application Programming Interface (API). This ensures a relatively high portability of the models as well as the interface, which can be executed on different instances. For example, the container can easily be started on a business platform to serve various use cases for end customers.

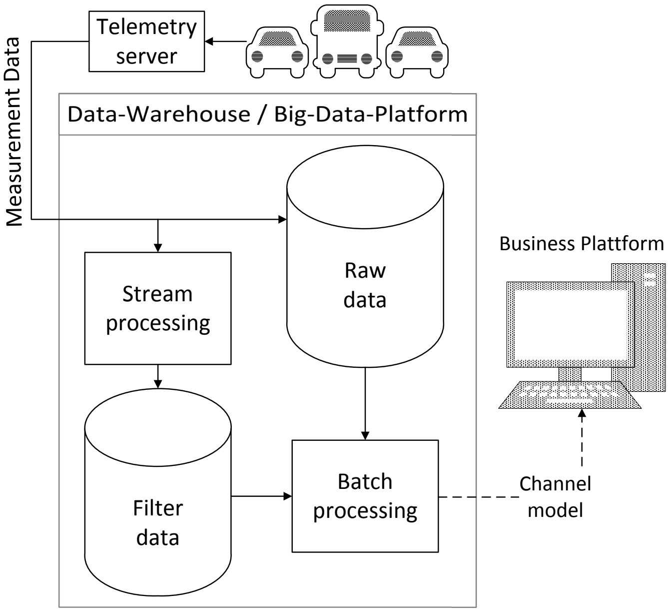


Fig. 2. Overview on the system architecture

D. Reverse Geocoding and Feature Engineering

In general, the term reverse geocoding (or inverse geocoding) describes searching and finding of mostly textual location information for existing coordinates. This process forms the basis for further data analysis with previously unknown attributes (”data mining”). This study primarily uses the open source database OSM as an additional source for geographically referenced data. Reverse geocoding functions include the determination of the OSM road type for the respective positions of transmitter and receiver vehicle, as well as the determination of the Line of Sight (LoS) or NLoS conditions. The determination of OSM road type is realized by adding the attribute of the OSM line element closest to the coordinate point. In total, nine different road types plus associated link types are considered and categorized as potentially passable by motor vehicles: *motorway, trunk, primary, secondary, tertiary, unclassified, residential, service, living_street* (see Fig. 3). The determination of the LoS/NLoS condition is based on geometric evaluation of whether the direct line of sight between both transceivers is intersected by an OSM polygon of the *building* type. If this is the case, an NLoS connection between transmitter and receiver vehicle is assumed, which is generally regarded as a damping influence on a radio connection. From a technical point of view, a database management system based on PostgreSQL with OSM database encapsulated in a docker container is used. The feature engineering is carried out by means of special SQL queries.

Fig. 3. Road type categories and buildings in OSM database

E. Propagation Modelling Approach

The channel modelling is based on a machine learning model. The model was trained with the recorded data of the vehicles (see III-B). The classical supervised learning approach needs features and labels. For this purpose, the euclidean distance between the vehicles (I_1), whether the line of sight was obscured by a building (I_2) and the road type the ego vehicle was driving on (I_3), are used as features. The OSM road types are grouped together in categories *major*, *minor* and *residential* to have sufficient data in every category. The features were encoded, normalized and scaled. The data was randomly shuffled, split into 80% training data and 20% test data. The recorded RSSIs between the vehicles are the labels for the model. The Tensorflow/Keras[1] Regression Network was designed as shown in Fig. 4 with two hidden layers. The output O_1 of the deep neural network is the predicted RSSI value for the ITS G5 communication in dBm.

The network architecture has sequential, densely connected layers with rectified linear unit activation functions as neurons. The mean square error between predicted RSSI and measured RSSI was minimized during training process with Adam optimizer. [7]

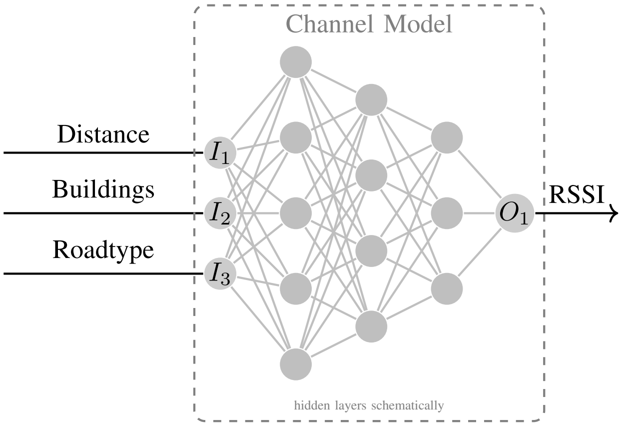
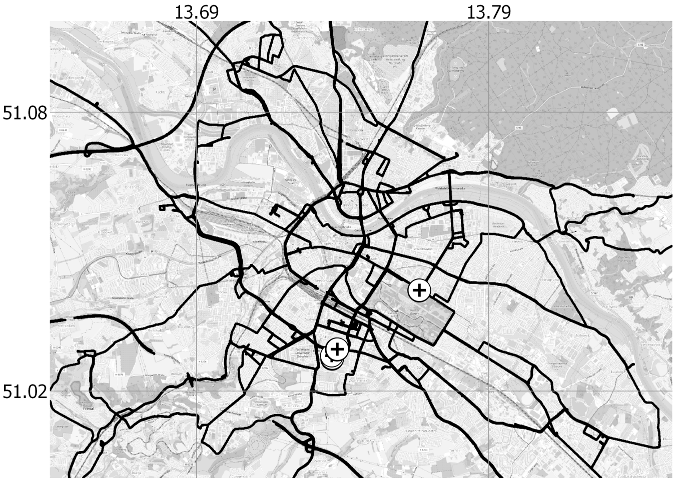


Fig. 4. Neural Network Architecture

IV. RESULTS

A. Data Collection

A sufficient amount of data is required to train the neural network that forms the basis of channel modeling. As this is an ongoing research project, the collection of data with the described system is still in progress. To date[2], the vehicles of the accident research unit, primarily recording V2V communication data, have sent approx. 5.0 million ETSI ITS G5 messages, of which the other vehicle has received approx. 3.3 million messages. The road network covered by previous trips in and around Dresden (Germany) is shown in Fig. 5.



basemap: ©OpenStreetMap contributors

Fig. 5. City of Dresden, Germany with Recorded Streets and RSUs (+)

As part of various initiatives to promote automated driving, ETSI ITS G5-capable RSUs are increasingly being installed and used in the city area for testing and demonstration purposes. Two of these stations, with which

---

[1] Tensorflow Version 1.12.0

[2] As of february 2019

V2X communication is already taking place in a real environment, are also marked in Fig. 5. This includes the rooftop installed RSU, specially added for the research project (see III-A). In order to obtain a certain statistical confidence for the empirical channel model, sufficient data should be available over the entire value range of the input data. At present, the data volume for communication connections with long distances between transceivers respectively a rather low RSSI is considered too insignificant. Thus, model predictions for long-distance communication are statistically uncertain.

B. Channel Model

The output of the neural network predicts an exponential distribution function of the received RSSI value as a function of distance, as shown in Fig. 6. This corresponds to the intuitive estimation and can be interpreted as an adapted free space path loss model (see Formula 2). In principle, the use of machine learning methods for the development of an empirical propagation model for V2X communication seems to be feasible.

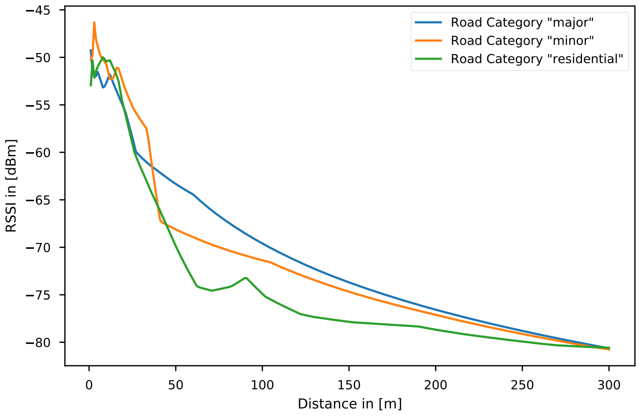


Fig. 6. Empirical Channel Model learned by a Neural Network for road categories with LoS-Conditions

C. Use Cases

An application pursued within the research project is the creation of a planning possibility for optimal positioning of ITS Roadside Stations. Embedded in an intuitive user interface, the use of the presented model compared to deterministic radio planning applications, e.g. using ray tracing methods, is characterized by its simple implementation and low resource consumption. This makes it particularly suitable for an initial estimation of the positioning of RSUs.

A second use case, which primarily uses the V2V propagation model, assesses potentials of ETSI ITS G5 with regard to the avoidance of traffic accidents and the reduction of accident severity. For this purpose, discretized driving trajectories of two accident participants of actual traffic accidents are evaluated retroactively with regard to the connection characteristics at and before the time of the collision..

V. CONCLUSION AND FUTURE WORK

The publication presents the basic design of a big data acquisition and processing system that uses V2X radio communication data from long-term measurements to generate a prototypical propagation model. The modeling is done utilizing a *Tensorflow* regression network. The possibility to enrich raw data with further characteristics was demonstrated exemplarily by the characteristics road type and NLoS. Necessary steps and the principle feasibility of creating an empirical V2X propagation model using machine learning methods are shown.

With the planned integration of a special measuring tram, a completely new type of measuring vehicle is also on the horizon. The evaluation of V2X radio communication data of a track-bound vehicle, whose local trajectory characteristics are completely identical due to the system, holds a high potential, especially with regard to position-independent influencing factors. On the side of potential applications, the possibility of an onboard prediction of the expected reception quality for a specific vehicle is an interesting option. For this purpose, the direct determination of the input vectors used for a concrete radio channel model by the vehicle itself would be necessary.

An obvious step to extend the existing system is to add further input features, which potentially influence the V2X communication. In addition to the already used features distance, road type and buildings, the extension by the following features is planned: tunnel environment, number of lanes, vehicle speed, interference by other transmitters, precipitation intensity, as well as time of day, month and year. While some of the mentioned variables act as direct influencing factors, e.g. additive reflections by tunnel walls, other variables are assumed to have an indirect effect on the propagation conditions. This applies, for example, to the time of day. During peak traffic hours, traffic density usually increases and so does the number of vehicles causing NLoS connections. Qualification and quantification of the individual influencing factors are of scientific relevance on the way to a more networked mobility.

REFERENCES

[1] T. Frankiewicz, A. Hinsberger, T. Lorenz, H.-J. Hilt, S. Weber, H. Wieker, and F. Köster, “Standortbestimmung und integration von its roadside stations für die anwendungsplattform intelligente mobilität,” *AAET-Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel*, vol. 9, no. 10.02, 2011.

[2] L. Cheng, B. E. Henty, D. Stancil, F. Bai, and P. Mudalige, “Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 ghz dedicated short range communication (dsrc) frequency band,” *Selected Areas in Communications, IEEE Journal on*, vol. 25, pp. 1501 – 1516, 11 2007.

[3] I. Sen and D. Matolak, “Vehiclevehicle channel models for the 5-ghz band,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, pp. 235–245, 06 2008.

[4] A. F. Molisch, F. Tufvesson, J. Karedal, and C. Mecklenbruker, “A survey on vehicle-to-vehicle propagation channels,” *Wireless Communications, IEEE*, vol. 16, pp. 12 – 22, 01 2010.

[5] D. Eckhoff, N. Sofra, and R. German, “A performance study of cooperative awareness in etsi its g5 and ieee wave,” in *2013 10th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, March 2013, pp. 196–200.

[6] Z. Hasani, M. Kon-Popovska, and G. Velinov, "Lambda architecture for real time big data analytic," *ICT Innovations*, pp. 133–143, 2014.

[7] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980