

#### Contents lists available at ScienceDirect

# Data in brief





## Data Article

# A dataset of full-stack ITS-G5 DSRC communications over licensed and unlicensed bands using a large-scale urban testbed



Andrea Tassi <sup>a, \*</sup>, Ioannis Mavromatis <sup>a</sup>, Robert J. Robert Piechocki <sup>a, b</sup>

### ARTICLE INFO

Article history: Received 2 April 2019 Received in revised form 17 July 2019 Accepted 29 July 2019 Available online 6 August 2019

Keywords: ITS-G5 DSRC RSSI V2V V2I V2X CAM

#### ABSTRACT

A dataset of measurements of ETSI ITS-G5 Dedicated Short Range Communications (DSRC) is presented. Our dataset consists of network interactions happening between two On-Board Units (OBUs) and four Road Side Units (RSUs). Each OBU was fitted onto a vehicle driven across the FLOURISH Test Track in Bristol, UK. Each RSU and OBU was equipped with two transceivers operating at different frequencies. During our experiments, each transceiver broadcasts Cooperative Awareness Messages (CAMs) over the licensed DSRC band, and over the unlicensed Industrial, Scientific, and Medical radio (ISM) bands 2.4 GHz-2.5 GHz and 5.725 GHz-5.875 GHz.

Each transmitted and received CAM is logged along with its Received Signal Strength Indicator (RSSI) value and accurate positioning information. The Media Access Control layer (MAC) layer Packet Delivery Rates (PDRs) and RSSI values are also empirically calculated across the whole length of the track for any transceiver. The dataset can be used to derive realistic approximations of the PDR as a function of RSSI values under urban environments and for both the DSRC and ISM bands — thus, the dataset is suitable to calibrate (simplified) physical layers of full-stack vehicular simulators where the MAC layer PDR is a direct function of the RSSI. The dataset is not intended to be used for signal propagation modelling.

<sup>&</sup>lt;sup>a</sup> Department of Electric and Electronic Engineering, University of Bristol, UK

b The Alan Turing Institute, London, NW1 2DB, UK

<sup>\*</sup> Corresponding author. E-mail address: a.tassi@bristol.ac.uk (A. Tassi).

The dataset can be found at https://doi.org/10.5523/bris.eu-powp7h3jl525yxhm3521f57, and it has been analyzed in the following paper: I. Mavromatis, A. Tassi, and R. J. Piechocki, "Operating ITS-G5 DSRC over Unlicensed Bands: A City-Scale Performance Evaluation," IEEE PIMRC 2019. [Online]. Available: https://arxiv.org/abs/1904.00464.

© 2019 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

#### Specifications table

Subject	Computer Science
Specific subject area	Computer Networks and Communications
Type of data	The network data exchanged by each RSU and OBU is provided as multiple PCAP traces [1] and CSV files. For each experimental session, each transceiver onboard each RSU and OBU is associated with a PCAP trace and CSV file. For each transceiver pertaining to each RSU and OBU, we also included a processed version of the data saved as MATLAB tables [2].
How data were acquired	The Innovate UK-funded FLOURISH Test Track, which includes multiple RSUs and OBUs and spans a length of 5km including key roads in the centre of Bristol, UK.
Data format	Raw data format, as recorded by the RSUs and OBUs pertaining to the FLOURISH Test Track, in the form of PCAP traces and CSV files. The raw data are then filtered in order to remove not required information from the PCAP traces and the CSV files, such as the parts of an IEEE 802.11p frames encapsulating CAMs. Filtered data has been recorded as MATLAB tables.
Parameters for data collection	The data has been recorded by employing two vehicles driving around the whole length of the FLOURISH Test Track. Since the data set has been recorded across multiple days, we relied on different drivers driving for ~ 4 h per day (namely, ~ 2 h in the morning and ~ 2 h in the afternoon). Furthermore, drivers experienced different traffic conditions in the city — thus the time needed to drive the whole length of the track depended on the time of the day.
Description of data collection	Eight data recording sessions of ~ 2h each, where one vehi- cle was driven in a clockwise and the other in an anticlockwise fashion across the whole length of the FLOURISH Test Track. Two transmission frequencies were tested during each data recording session (one per transceiver).
Data source location	A typical urban setting in Bristol, UK, comprising the following roads: Marlborough Street, Upper Maudlin Street, Park Row, Woodland Road and Queens Road.
Data accessibility	Data is available online at the link below: https://doi.org/10.5523/bris.eupowp7h3jl525yxhm3521f57
Related research article	I. Mavromatis, A. Tassi and R. J. Piechocki, "Operating ITS- G5 DSRC over Unlicensed Bands: A City-Scale Performance Evaluation," IEEE PIMRC 2019. [Online]. Available: https://arxiv.org/abs/1904.00464.

#### Value of the data

- Our dataset consists of network interactions recorded among RSUs and OBUs in an urban environment in both a Vehicle-to-Infrastructure (V21) and Vehicle-to-Vehicle (V2V) fashion. Network interactions account for CAMs exchanged according to the standard ITS-G5 DSRC, operated over the licensed DSRC band and unlicensed ISM bands 2.4 GHz-2.5 GHz and 5.725 GHz-5.875 GHz. Along with the dataset we also provided the necessary data processing tools to estimate the MAC layer PDR as a function of RSSI values. The empirical PDR value can be directly integrated into (simplified) physical layers that are widely adopted in full-stack network simulators. It is beyond the scope of our experiments to perform a measurement campaign suitable for deriving accurate propagation models.
- By building upon the empirical PDR values (expressed as a function of the RSSI values), experts from the Wireless
  Communications and Networks communities can directly assess the impact of operating an ITS-G5 DSRC system over both
  licensed and unlicensed frequency bands, being respectively the DSRC and the ISM bands. For instance, this will facilitate
  further research on the provision of safety-critical and non-safety critical automotive services over the DSRC band, the ISM
  bands, or both at the same time.
- Licensed DSRC band is expected to be (at least partially) allocated to Cellular-V2X (C-V2X) systems. As such, the empirical PDR values (expressed as a function of RSSI values) derived from this dataset can be directly compared against empirical

PDR values measured for C-V2X systems, in urban environments. Furthermore, a similar performance investigation can be carried out by comparing the empirical PDR values of C-V2X against ITS-G5 DSRC operated over the unlicensed ISM bands we considered in our experiments.

• Since our dataset has been recorded employing vehicles driving in a real urban environment; it can also be used by cyber-security experts. A potential usage is to establish what the "standard behavior" of a connected vehicle is when observed from the communication perspective — thus making it possible to train anomaly detection systems to identify potential threats [3]. For instance, a vehicle can maliciously broadcast CAMs where its location has been altered. However, any RSU receiving malicious CAMs will be able to measure their RSSI value. Since the RSSI value is a function of the environment and the distance between the transmitter and the receiver, an anomaly detection system can verify if a measured RSSI value is comparable with the expected value given the positioning information encapsulated into each CAM.

#### 1. Data

The raw data associated with each transceiver (namely, PCAP traces [1] and CSV files) installed on each RSU and OBU are organized as follows. During each day of trial, our dataset includes one or two sequences of raw data files for each transceiver. That depends on if the transceiver is installed on either an RSU or an OBU. In particular, raw data files associated with each RSU collates two experimental sessions (namely, morning and afternoon) per-file, while raw data files pertaining to each OBU always refer to a single experimental session.

As for the PCAP traces, each of them records the whole of the network interaction where each transceiver is involved in. In particular, each received IEEE 802.11p frame is recorded into a PCAP trace with its corresponding RSSI value. Our PCAP traces can be accessed using our modified version of the tool Tcpdump [4]. Furthermore, for each transceiver, key fields of each of the transmitted and received CAMs are recorded in a tabular format in the same CSV file. Since, the FLOURISH Testing Track logs both the transmitted and received CAMs in the same raw data file, for the sake of data readability and debugging purposes, the labels of each column are replicated in each data row, i.e., each data row follows the format:

Column\_Label\_0; Value\_0; Column\_Label\_1; Value\_1; ... Column\_Label\_n; Value\_n

For the sake of reducing the overall size of the dataset and hence, making it possible to manipulate the dataset with a smaller memory footprint, the raw data has been filtered into MATLAB data files [2]. In particular, for each set of CSV file and PCAP trace associated to a given transceiver, we generated the corresponding MATLAB tables for the transmitted/received CAMs, and one MATLAB table per-PCAP trace referring to the received CAMs. Further details about the labels of the columns in the CSV raw data files and MATLAB tables are provided in Table 1.

The MATLAB framework designed for the above data manipulation and filtering can be found online in [5]. Our framework provides also tools to investigate the Key Performance Indicators (KPIs) shown in Figs. 3–10. These figures will be further discussed in the next section.

## 2. Experimental design, materials, and methods

The four RSUs providing network coverage across the 5 km length of the FLOURISH Testing Track (see Fig. 1) are fitted outside premises of the University of Bristol. Each site, along with its ID and characteristics, is listed below:

- MVB Merchant Ventures Building, BS8 1UB, UK: RSU installed at a height of ~8 m opposite to a Tiunction.
- DH Dorothy Hodgkin Building, BS1 3NY, UK: RSU located at ~12 m from the ground beside a curvy road.
- HW Hawthorns Building, BS8 1UQ, UK: RSU mounted at ~5 m from street-level, some foliage on site.
- SU Students Union Building, BS8 1LN, UK: RSU fitted at ~25 m from the ground beside a straight stretch of road.

**Table 1**Definition of the labels of each column of the CSV raw data files and MATLAB tables in the dataset.

Field Name	Definition	TX Entry	RX Entry
TX-REQ-CAM	REQ-CAM Transmitted CAM		
RxMAC	MAC address of the transmitter		/
RX-REQ-CAM	Received CAM		/
Protocol	ITS-G5 protocol ID [6]	✓	
Validation	ITS-G5 validation ID [6]		1
StationID	ID of the transmitter/receiver	/	/
GenDeltaTime	The remainder of Timestamp/65546 [6]	/	/
SeqNum	Sequence number of the transmitted/received CAM [7]	/	/
GpsLon	Longitude of the current position of the transmitter/receiver	/	/
GpsLat	Latitude of the current position of the transmitter/receiver	/	/
CamLon	Longitude of the position of the transmitter/receiver as encapsulated in a CAM	✓	1
CamLat	Latitude of the position of the transmitter/receiver as encapsulated in a CAM	✓	✓
SpeedConf	Encoding configuration for the speed [6]		✓
VehHeading	Heading of the transmitter		/
GpsSpeed	Speed of the transmitter/receiver	/	/
CamSpeed	Speed of the transmitter as encapsulated in a CAM	/	/
Timestamp	Timestamp at the moment of creation of a CAM	/	/
CamLength	Byte-length of a CAM	1	
RSSI	The RSSI value of a received CAM		✓

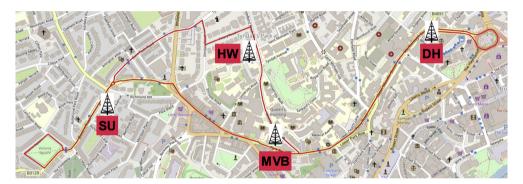
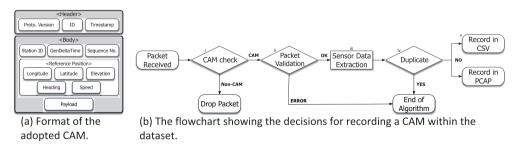


Fig. 1. The Flourish Testing Track and the locations of the four RSUs.



**Fig. 2.** Considered CAM and flowchart of the data logging process. When a packet is received in the TX or the RX queue, it is checked whether it is CAM message at first. Later, the validity of the packet is specified before being recorded in our dataset.

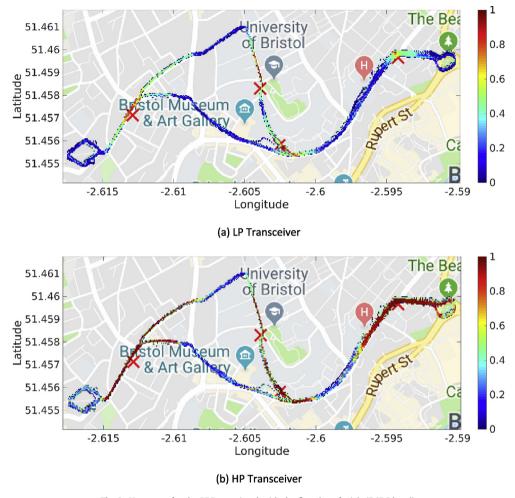


Fig. 3. Heatmaps for the PDR associated with the first day of trials (DSRC band).

All aforementioned sites host an identical RSU operating two transceivers at the same time. Since the first transceiver has a maximum transmission power of 25dBm and the second can reach up to 29dBm, in our dataset, we regard the first one as the Low Power (LP) transceiver and the second one as the High Power (HP) transceiver.

During each experimental session, two vehicles have also been used, with the only exception of the afternoon experimental session held on 15th February 2019 when only one vehicle had been used. When two vehicles were present, one was driven in a clockwise direction while the other in an anticlockwise fashion. Each vehicle was equipped with an OBU operating the same set of transceivers that we employed in each RSU. Further details about both the hardware and software pertaining to the RSUs and OBUs can be found in [8]. During each day of trials, the LP and HP transceivers onboard each RSU and OBU have been operated on 10 MHz-width channels. The center frequencies chosen for each day, the considered trans-mission power and Modulation and Coding Scheme (MCS) are listed in Table 2. Since the data traffic exchanged between the OBUs and RSUs was entirely made of BTP and CAM messages, we adopted the MCS QPSK-1/2 as imposed by the ITS-G5 DSRC standard [9].

All the RSUs and OBUs communicate using ETSI's ITS-G5 standard [10] and broadcast a CAM every 10 ms. Our setup is built upon the IEEE 802.11p DSRC physical layer as described before. On top of that, a

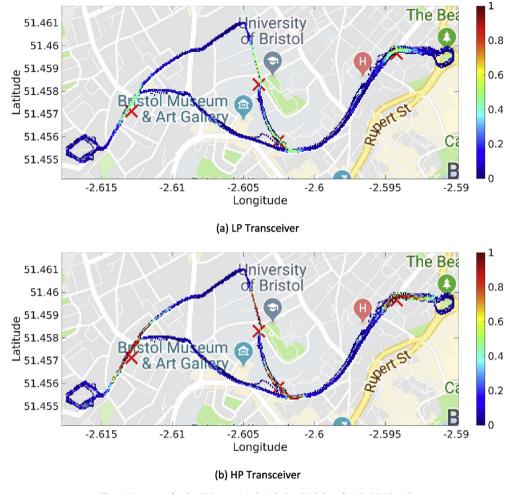


Fig. 4. Heatmaps for the PDR associated with the third day of trials (ISM band).

MAC and Logical Link Control (LLC) layers act as the adaptation between the physical and the upper layers. Later, the Basic Transport Protocol (BTP) operating on top of the MAC and the LLC, provides an end-to-end connection-less transport service in an ad-hoc fashion. The BTP is also responsible for multiplexing/de-multiplexing packets originating from/to the Facilities layer [10]. The exchanged CAMs are generated at the level of the Facilities layer. During our trials, we only referred to CAMs carrying positioning information (namely, position, speed and heading) pertaining to each OBUs and RSUs — obviously, in the case of the RSUs, both speed and heading is set to zero. For our research activities, we extended the standard CAM format described in [6] by adding two extra fields, namely, the Timestamp and the Sequence number. The usage of these fields has been described in [8], and it allowed us to reconcile the logged data during the data processing phase. All the fields encapsulated within the exchanged CAMs can be seen in Fig. 2(a).

Each transmitted and received CAM has been recorded both in the CSV and PCAP traces. When a packet appears in the transmission or reception queue of a transceiver, a decision is made whether it is going to be recorded or not. The flowchart describing that, can be seen in Fig. 2(b). For the sake of limiting the size of the dataset, only the CAMs exchanged were recorded in the traces (Fig. 2(b)-i). The rest of the wireless networks interactions within our nodes are discarded. The validity of CAMs within

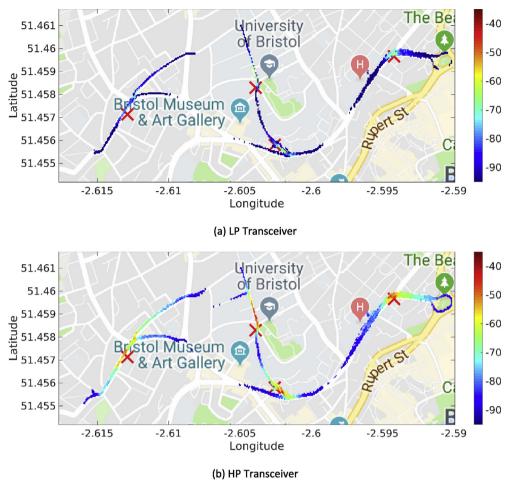


Fig. 5. Heatmaps for the RSSI values (expressed in dBm) associated with the first day of trials (DSRC band).

the queues is then checked (Fig. 2(b)-ii) by evaluating the length of the packet, and the expected fields based on the legacy CAM structure. For each valid CAM, the positioning information is extracted (Fig. 2(b)-iii). Subsequently, duplicated packets are then removed from the transmission/reception queue (Fig. 2(b)-iv). Whenever, the CAM is not valid or duplicates, the messaged is not recorded in the log files. Otherwise, the message is logged (Fig. 2(b)-v).

In particular, for what concerns an RSU, raw data recorded during a specific day of trial and pertaining to a specific transceiver, is recorded under the following filename and directory naming convention:

./rawData/ddmmyyyy/RSUs-ddmmyyyy/log\_XX\_YY.txt and

./rawData/ddmmyyyy/RSUs-ddmmyyyy/log\_XX\_TCP\_YY.pcap where "ddmmyyyy" signifies the date when the experiment took place, "XX" can either be "HP" or "LP" depending on the transceiver under consideration, and "YY" refers to one of the four RSUs (namely, "MVB", "DH", "HW" or "SU"). Thus, for instance, files "./rawData/11022019/RSUs-11022019/log\_HP\_MVB.txt" and "./rawData/11022019/RSUs- 11022019/log\_HP\_TCP\_MVB.txt" refer to the HP transceiver of the RSU located on the MVB site during the 11th February 2019. A similar conversion has been adopted for recording raw data associated with the OBUs, which is defined as follows:

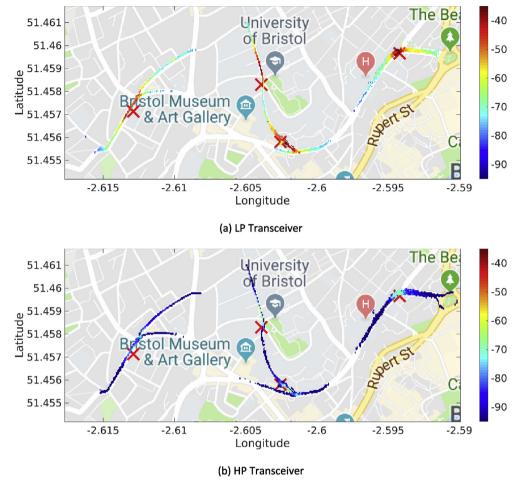


Fig. 6. Heatmaps for the RSSI (expressed in dBm) associated with the fourth day of trials (ISM band).

./rawData/ddmmyyyy/OBU-ZZ-ddmmyyyy-TT/log\_XX.txt and

./rawData/ddmmyyyy/OBU-ZZ-ddmmyyyy-TT/log\_XX\_tcpd.pcap where "ZZ" can be either equal to "00" or "01" depending of the fact the data refers to the first or the second OBU, respectively. Then, if data refers to an experimental session that took place in the morning then "TT" is equal to "Morning", otherwise, "TT" is set equal to "Afternoon". For instance, files "./rawData/11022019/OBU-00-11022019-Morning/log\_HP.txt" and "./rawData/11022019/OBU-00-11022019-Morning/log\_HP\_tcpd.pcap" refer to the experimental session recorded in the morning of 11th February 2019 and referring to the HP transceiver of the first OBU.

The raw data part of our dataset has been filtered by means of our MATLAB framework [5]. The output of the data filtering operations has been made available under the directories "importedData/ddmmyyyy", where files follow the naming convention summarized in Table 3.

Our MATLAB framework [5] can also be used to investigate the following Key Performance Indicators (KPIs). For simplicity, a representative subset of figures was chosen for each KPI:

• Heatmaps for the PDR values associated with the HP/LP transceivers across the roads forming the FLOURISH Testing Track. A sample of the sixteen resulting figures is shown in Figs. 3—6. A CAM is

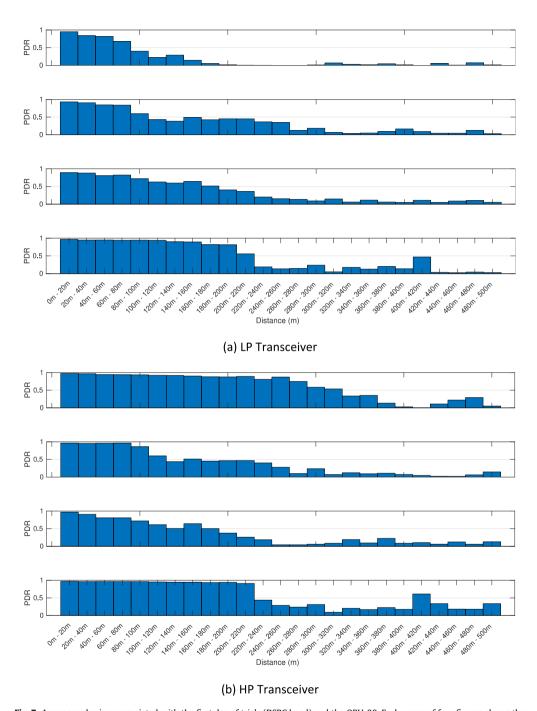


Fig. 7. Awareness horizon associated with the first day of trials (DSRC band) and the OBU-00. Each group of four figures shows the awareness horizon associated with the RSU at sites MVB, DH, HW and SU.

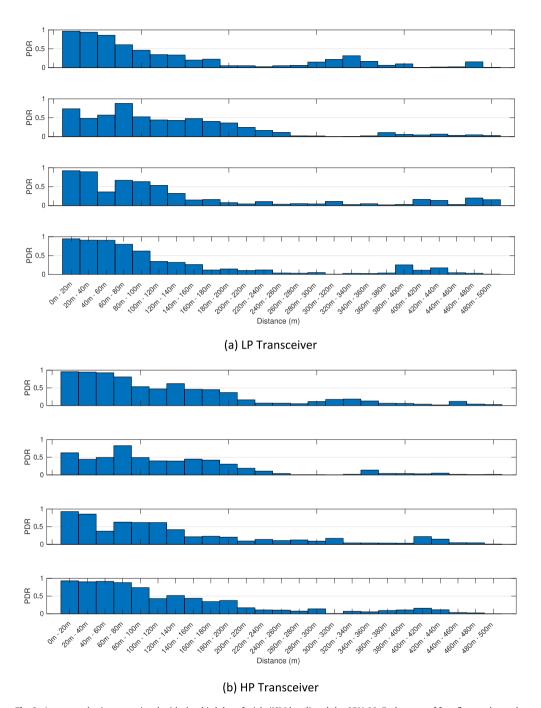


Fig. 8. Awareness horizon associated with the third day of trials (ISM band) and the OBU-00. Each group of four figures shows the awareness horizon associated with the RSU at sites MVB, DH, HW and SU.

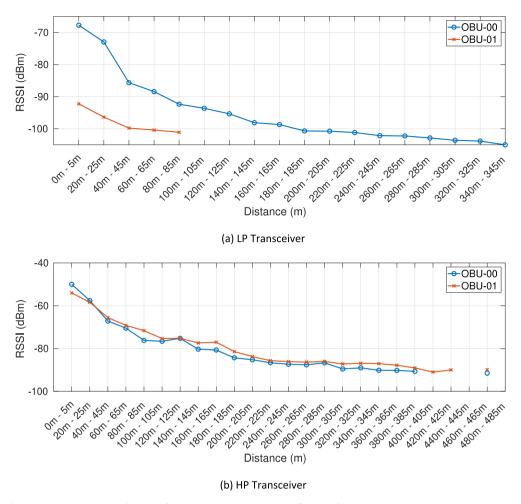
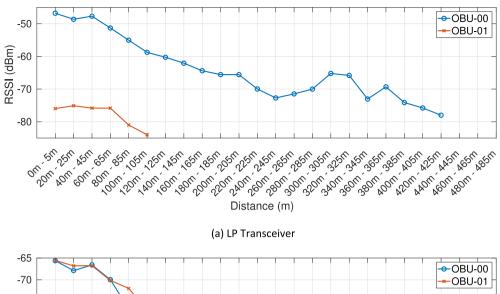


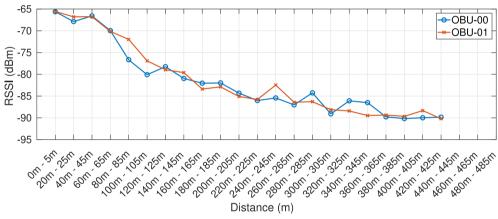
Fig. 9. RSSI measurement as a function of the distance, associated with the first day of trials (DSRC band) and the RSU at the DH site.

considered as successfully delivered, when all the en-capsulated sensor information can be successfully extracted (as shown in Fig. 2(b)).

- Awareness horizons for both HP and LP transceivers showing the PDR as a function of the distance
  to a given RSU. A sample of the sixteen resulting figures is shown in Figs. 7 and 8. The successful
  CAM reception is considered as in the heatmaps KPI.
- RSSI value for both HP and LP transceivers as a function of the distance to a given RSU. A sample of the thirty-two resulting figures is shown in Figs. 9 and 10.
- Our framework can be used to generate the entire set of figures for all the above-mentioned KPIs. It can also be further extended to accommodate different or more specific KPIs per-taining to the network interactions at specific road junctions at specific time of the day, or the network interactions among moving vehicles. As for the latter, for convenience, our MATLAB framework [5] includes a complementary script (namely, "v2vInteractions/v2vTimings.m") to locate the timestamps when each vehicle received a CAM from another one, and saves them into MATLAB table.

Our MATLAB framework also gives the reader the option of filtering the entire dataset to retrieve all the network interactions between the pair of OBUs considered in our experiments. In particular, when





(b) HP Transceiver

Fig. 10. RSSI measurement as a function of the distance, associated with the fourth day of trials (ISM band) and the RSU at the HW site.

**Table 2**The center frequency, transmission power and MCS used throughout the four days of trials.

Date	HP Transceivers	LP Transceivers	Band	HP Power	LP Power	MCS [9]
1st Day - 11/02/2019	5.900 GHz	5.890 GHz	DSRC	29 dBm	25 dBm	QPSK 1/2
2nd Day - 13/02/2019	5.200 GHz	2.437 GHz	ISM			
3rd Day - 15/02/2019	5.180 GHz	2.412 GHz	ISM			
4th Day - 18/02/2019	5.320 GHz	2.462 GHz	ISM			

**Table 3**Naming conventions and per-file content description for the filtered part of our dataset.

File Name	Content
rsu_YY_TxCAM_XX.mat	Data from CAMs transmitted by the RSU at site "YY" (namely, "MVB", "DH", "HW" or "SU"), from the transceiver "XX" (namely, "LP", or "HP")
rsu_YY_RxCAM_XX.mat	Data from CAMs received by the RSU at site "YY" (namely, "MVB", "DH", "HW" or "SU"), from the transceiver "XX" (namely, "LP", or "HP")
rsu_YY_tcpDump_XX.mat	Data originating from the PCAP traces associated with the RSU at site "YY" (namely, "MVB", "DH", "HW" or "SU"), from the transceiver "XX" (namely, "LP", or "HP")
vehNo_ZZ_TxCAM_XX_UU.mat	Data from CAMs transmitted by the OBU at site "ZZ" (namely, "00" or "01"), from the transceiver "XX" (namely, "LP", or "HP"), during the "UU" session (namely, morning "mr" or afternoon "af")
vehNo_ZZ_RxCAM_XX_UU.mat	Data from CAMs received by the OBU at site "ZZ" (namely, "00" or "01"), from the transceiver "XX" (namely, "LP", or "HP"), during the "UU" session (namely, morning "mr" or af- ternoon "af")
vehNo_ZZ_tcpDump_XX_UU.mat	Data originating from the PCAP traces associated with the OBU at site "ZZ" (namely, "00" or "01"), from the transceiver "XX" (namely, "LP", or "HP"), during the "UU" session (namely, morning "mr" or afternoon "af")

the script "v2vInteractions/v2vInteractions.m" is executed, it produces a table-formatted MAT file (for each day and transceiver) defined by the columns summarized in Table 1 where each row corresponds to a transmitted/received CAM. For the fields that exist on both the TX and the RX sides (e.g. the "GpsLon"), we distinguish their values in the table by adding either "-RX" or "-TX" at the column name (e.g. "GpsLon-TX" and "GpsLon-RX"). Similarly, the reader can develop a filtering script to retrieve the network interactions pertaining to a given pair of OBU and RSU. The starting point can be any of the existing scripts located in the folder "scriptResults" that can be modified as follows:

- 1. Modifying the two main loops of the script, namely, the loops associated with the day and the transmission power. The user can then choose to evaluate the results for a specific day or TX power. Based on that, our framework will load only the network interactions related to the chosen day and TX power.
- 2. Modifying the internal loops of our script, namely, the ones related to the OBUs and the RSUs. One or more specific pairs of OBU/RSU can be chosen to be processed.
- 3. Similarly to the "v2vInteractions.m" script, the unique sequence number for the transmitted packets per session, as well as the MAC address of the transmitter can be parsed at first.
- 4. Using the above information, it can be identified whether a packet is successfully delivered or not, i.e., checking whether a packet with the given MAC address and sequence number exists in the list of received packets.
- 5. Having this pairwise correlation between transmitted and received packets, a table can be later generated by incorporating all the given interactions between a set of devices (as in "v2vInteractions.m").
- 6. The above steps can be replicated for the requested number of OBU-RSU pairs generating a list of all the desired network interactions.

# Acknowledgments

This work is part of the FLOURISH Project, which is supported by Innovate UK, under Grant 102582.

<sup>&</sup>lt;sup>1</sup> For simplicity we focused on the position of the vehicles (both the one from the GPS receiver and the one encapsulated in each CAM), the TX timestamp and the MAC addresses of the OBUs. The rest (such as the "VehHeading" for example), can be easily added by extending the "v2vInteractions/v2vInteractions.m" script and including the new columns into the table.

#### Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### References

- [1] Libpcap, The tcpdump group, Online, Tech. Rep. (2018). Available: https://github.com/the-tcpdump-group/libpcap.
- [2] MAT-File Format, MathWorks, Online, Tech. Rep. (2018). Available: https://www.mathworks.com/help/pdf\_doc/matlab/matfile\_format.pdf.
- [3] C.S. Pasareanu, D. Gopinath, H. Yu, Compositional Verification for Autonomous Systems with Deep Learning Components, Springer International Publishing, Cham, 2019, pp. 187–197.
- [4] Online, Modified Version of Tcpdump, University of Bristol, Tech. Rep., 2019. Available: https://github.com/v2x-dev/tcpdump.
- [5] Online, ITS-G5 Trial Processing Scripts, University of Bristol, Tech. Rep., 2019. Available: https://github.com/v2x-dev/ITS-G5\_trials.
- [6] Specification of cooperative awareness basic service EN 302 637-2, ETSI, Tech. Rep., Nov. (2014).
- [7] A. Tassi, I. Mavromatis, R. Piechocki, A. Nix, C. Compton, T. Poole, W. Schuster, Agile data offloading over novel fog computing infrastructure for CAVs, in: Proc. Of IEEE VTC 2019-Spring, Kuala Lumpur, ML, Apr, 2019.
- [8] I. Mavromatis, A. Tassi, R. Piechocki, Operating ITS-G5 DSRC over Unlicensed Bands: A City-Scale Performance Evaluatio, Online, IEEE PIMRC, 2019. Available: https://arxiv.org/abs/1904.00464.
- [9] STDMA recommended parameters and settings for cooperative ITS; access layer part TR 102 861, ETSI, Tech. Rep. (Jan. 2011).
- [10] Access layer specification for intelligent transport systems operating in the 5 GHz frequency band EN 302 663, ETSI, Tech. Rep. (May 2013).