AI4MOBILE

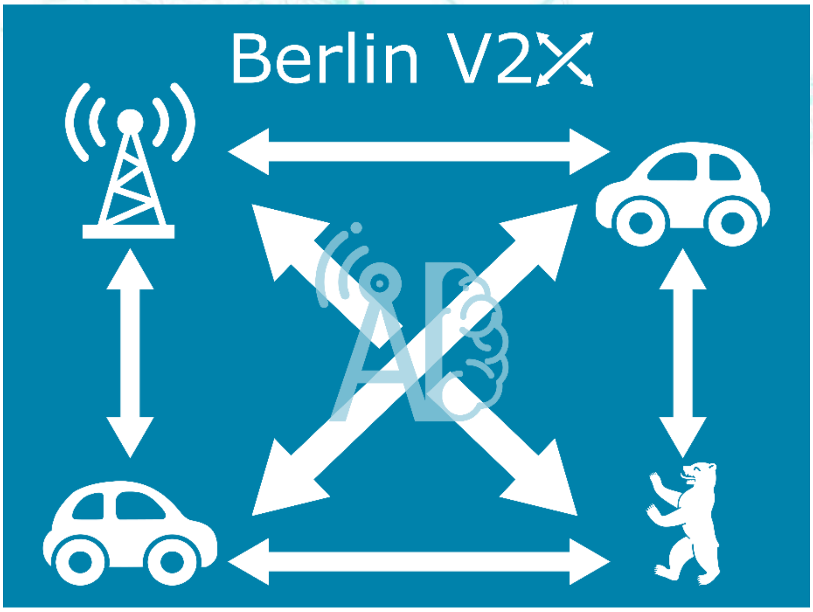
Federal Ministry of Education and Research

BMW GROUP · MINI · ROLLS-ROYCE · Deutsche Telekom · ERICSSON · Fraunhofer HHI · Technische Universität Berlin · Technische Universität Dresden · Technische Universität Kaiserslautern · vodafone

Berlin V2X

A Machine Learning Dataset from Multiple Vehicles and Radio Access Technologies

Instructions



Contents

Project Details

Call	*Artificial Intelligence in Communication Networks*
Project Coordinator	*Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute* *Prof. Dr.-Ing. Slawomir Stanczak*
Project start date	*15th of March 2020*
Duration	*36 months*

Berlin V2X

Download the dataset on IEEE Dataport

Check the code and documentation on GitHub

The Berlin V2X dataset offers high-resolution GPS-located wireless measurements across diverse urban environments in the city of Berlin for both cellular and sidelink radio access technologies, acquired with up to 4 cars over 3 days. The data enables thus a variety of different machine learning (ML) studies towards vehicle-to-anything (V2X) communication.

In the following, an overview of the data is provided. For a detailed description of the measurement campaign please refer to the paper.

Requirements

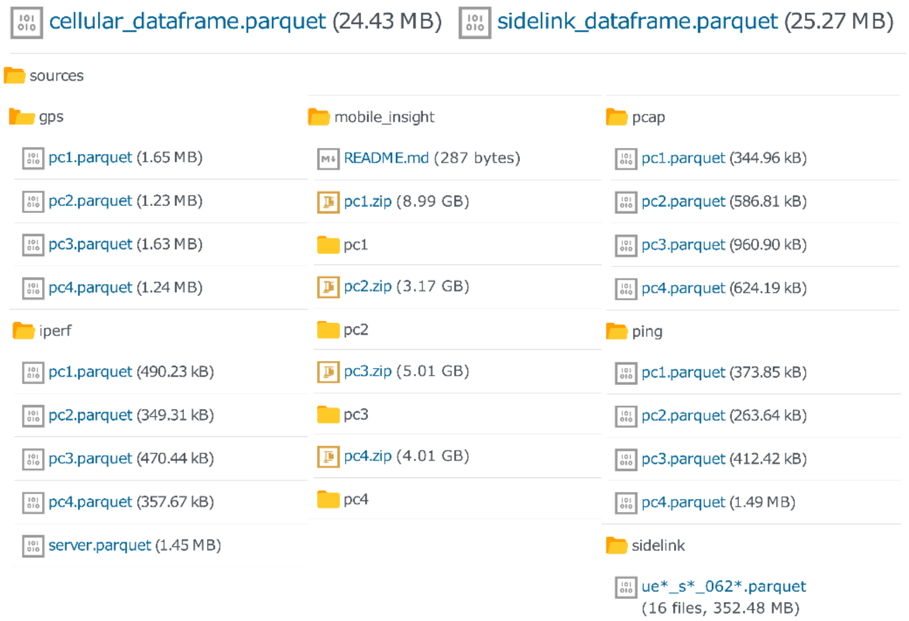
We strongly recommend to work on Python with the following libraries:

- pandas
- pyarrow

Furthermore, we suggest some additional libraries to process and analyze the data, such as:

- numpy for common mathematical tools
- jupyter to run the interactive examples
- seaborn and matplotlib for plotting
- folium and related (branca, smopy) for map visualization
- scikit-learn for ML analysis

File overview



cellular_dataframe.parquet (24.43 MB)
sidelink_dataframe.parquet (25.27 MB)

sources

gps
pc1.parquet (1.65 MB)
pc2.parquet (1.23 MB)
pc3.parquet (1.63 MB)
pc4.parquet (1.24 MB)

iperf
pc1.parquet (490.23 kB)
pc2.parquet (349.31 kB)
pc3.parquet (470.44 kB)
pc4.parquet (357.67 kB)
server.parquet (1.45 MB)

mobile_insight
README.md (287 bytes)
pc1.zip (8.99 GB)
pc1
pc2.zip (3.17 GB)
pc2
pc3.zip (5.01 GB)
pc3
pc4.zip (4.01 GB)
pc4

pcap
pc1.parquet (344.96 kB)
pc2.parquet (586.81 kB)
pc3.parquet (960.90 kB)
pc4.parquet (624.19 kB)

ping
pc1.parquet (373.85 kB)
pc2.parquet (263.64 kB)
pc3.parquet (412.42 kB)
pc4.parquet (1.49 MB)

sidelink
ue*_s*_062*.parquet
(16 files, 352.48 MB)

Berlin V2X includes different data files in parquet format. For an easy on-boarding, we provide GPS-located and labelled data frames, merged and resampled to 1 second, for:

- Cellular measurements (*cellular_dataframe.parquet*).
- Sidelink measurements (*sidelink_dataframe.parquet*).

All data sources are also provided in high-resolution for:

- Ping traces - per car (x4).
- Iperf traces - per car (x4) for downlink measurements + server (x1) for uplink measurements.
- TCPdump data for the sidelink - per car (x4).
- GPS data - per car (x4).
- MobileInsight traces - per car (x4) and message type (x50).
    - The complete messages are zipped per device. The following message types, which have been merged into the cellular dataframe, are also provided unzipped for a quick access:
        - LTE_PHY_Serv_Cell_Measurement
        - LTE_RRC_Serv_Cell_Info
        - LTE_PHY_PDSCH_Stat_Indication
        - LTE_PHY_PUSCH_Tx_Report

Data category	Source	Tool	Sampling interval	Features
Cellular	DME	Mobile Insight	10 ms	PHY: SNR, RSRP, RSRQ, RSSI
			20 ms	PDSCH/PUSCH: RBs, TB Size, DL MCS, UL Tx Power
			Event-based	RRC: Cell Identity, DL/UL frequency, DL/UL bandwidth
		ping	Event-based	Jitter, delay
		iperf	1 s	DL datarate
	Server	iperf	1 s	UL Datarate
Sidelink	SDR UE	tcpdump	Event-based	SNR, RSRP, RSRQ, RSSI, Noise Power, Rx Power, Rx Gain
Position	GPS		1 s	Latitude, Longitude, Altitude, Velocity, Heading
Side information	Internet database	HERE API	5 min	Traffic Jam Factor, Traffic Street Name, Traffic Distance
		DarkSky	1 hour	Cloud cover, Humidity, Precipitation Intensity & Probability, Temperature, Pressure, Wind Speed

For merging and preprocessing details check the notebooks in the *preprocess* folder.

Quickstart

You can directly load the cellular or sidelink merged dataframe in pandas and inspect the columns. Some general information on each column can be found in *sidelink_info.csv* and *cellular_info.csv*.

In order to reuse the code in *analyze* and *preprocess*, place the dataset under *data*. The publication figures will be saved to *plots*.

Data sources

1. MobileInsight
2. Iperf
3. Ping
4. RUDE & CRUDE
5. TCP Dump
6. GPS traces
7. Here API
8. DarkSky API

MobileInsight

All information captured by MobileInsight from available LTE channels. The available information also depends on the modem of the measurement device.

Information from the following message types is provided in the merged cellular dataframe:

- LTE_PHY_Serv_Cell_Measurement
- LTE_RRC_Serv_Cell_Info
- LTE_PHY_PDSCH_Stat_Indication
- LTE_PHY_PUSCH_Tx_Report

The merged data is based in the signal strength and quality PHY Serving Cell information for both primary and secondary cells. This has been enriched with RRC information on the cells and the aggregated number of allocated resource blocks and transport block size in downlink and uplink from PDSCH/PUSCH, respectively. The shared channel information also allows us to include the transmitted power in uplink and the modulation and coding scheme (MCS) in downlink.

More details about the preprocessing of MobileInsight data can be found under the *mi* folder.

Iperf

Iperf is a speed test application that enables measuring the bandwidth and jitter of a UDP or TCP connection.

In the measurement campaign, Iperf was run on both a DME and server to receive throughput measurements with a granularity of **1s**. For experiments that require high accuracy. The iperf measurements that have been merged in the cellular dataframe are extracted from the destination, i.e., DME for downlink and server for uplink.

Ping

Collected from the console command `ping`, it provides the delay measurements.

RUDE

The packets were transmitted according to scenarios S1 and S2:

- **S1**: Cooperative awareness messages (CAM) of length 69 bytes at a packet transmission rate of 20 Hz and modulation and coding scheme (MCS) 8 using two sub-channels.
- **S2**: Collective perception messages (CPM) of length 1000 bytes (including IP header and payload) at a rate of 50 Hz with MCS 12 using ten sub-channels.

We aggregate the information on the received packets down to 1 second to estimate packet error rate. For a detailed insight of the sidelink signal parameters, check the dataset publication or RUDE's documentation.

TCP Dump

TCP Dump is a packet analyzer that allows tracking transmitted packets and their properties (e.g. payload, size of the packet). The received sidelink packages were decoded with TCPdump and parsed into parquet files for every receiver. The sidelink data extracted from the incoming messages for any given sidelink UE are provided as separate parquet files.

GPS

GPS data is collected for each device with a granularity of 1 second.

The GPS traces are enriched with the variable `Pos in Ref Round`, i.e., the position in the reference round. This variable is a mapping of the latitude and longitude into the distance that was driven by an arbitrarily chosen car from an arbitrarily chosen point in an arbitrarily chosen round. In this way, `Pos in Ref Round` allows the analysis of wireless parameters against a single 1-dimensional spatial value (Check add_pos_in_ref_round.ipynb for details). For the sidelink data, the distance between cars is also provided.

Distance is computed in all cases after conversion to planar coordinates for simplicity. The error should be negligible due to the small area that is covered by the data.

The merged GPS traces in the IEEE dataport also include the side information from the APIs HERE and DarkSky.

HERE API

The information about traffic density was downloaded from the HERE Traffic API every 5 minutes during the measurements.

For determining the Traffic Jam Factor at a given location the closest route from the API data was calculated. For debugging purposes the name of the street where this Traffic Jam Factor was taken from is saved in the Traffic Street Name column and the distance to this street is saved in the Traffic Distance column.

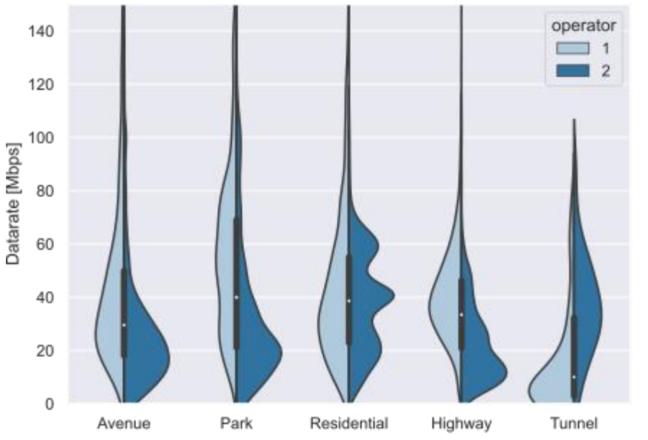
DarkSky

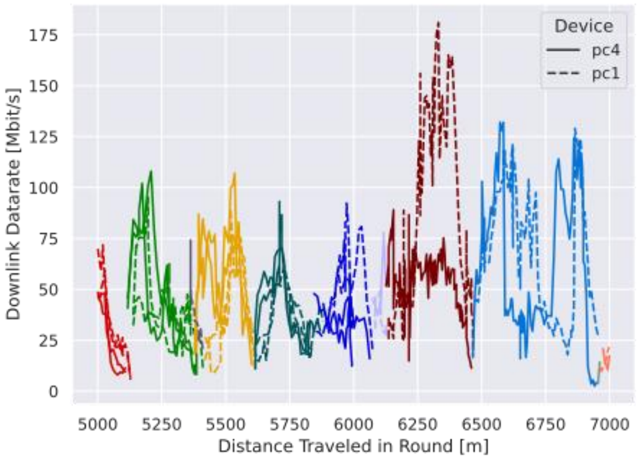
The information about Cloud cover, Humidity, Precipitation Intensity & Probability, Temperature, Pressure and Wind Speed was downloaded from the DarkSky API.

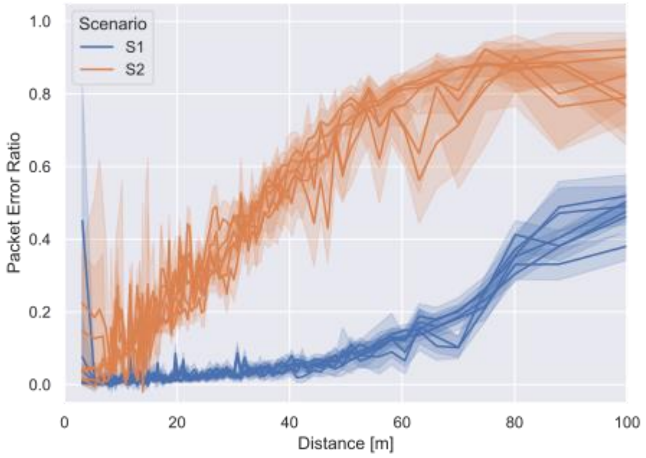
Time granularity is 1 hour and location granularity is 0.01 degrees in both latitude and longitude.

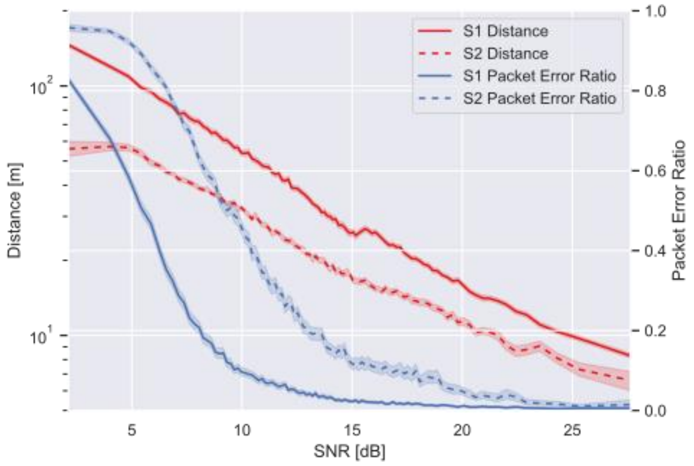
Reference

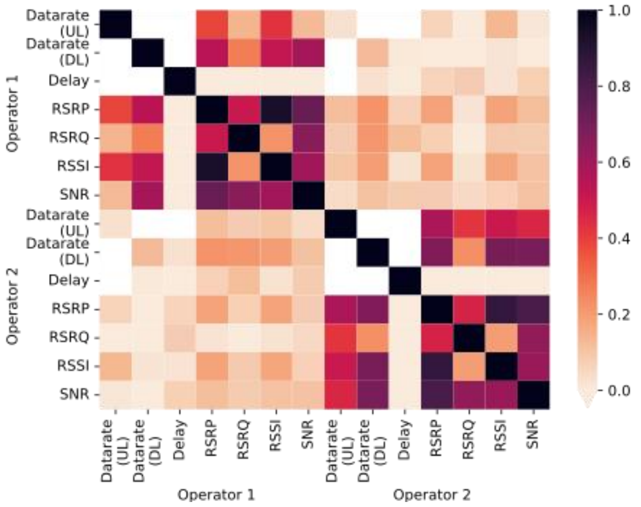
Examples

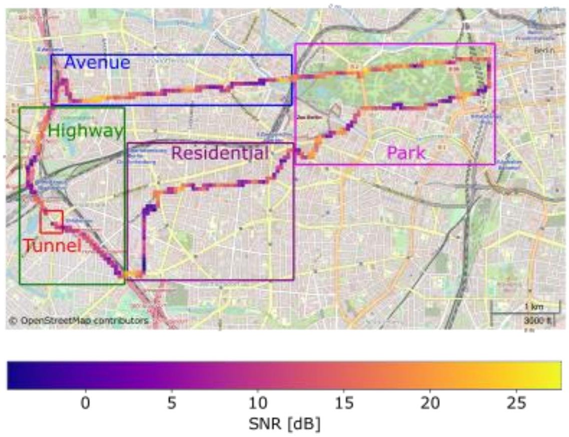












The code to generate the publication figures can be found in the *analyze* folder.

AI4Mobile

AI4Mobile is a research project funded by the Federal Ministry for Education and Research (BMBF), from the announcement Artificial Intelligence in Communication Networks within the scope of the High-Tech Strategy of the German Federal Government.

The scope of the project is the study of AI-aided wireless systems for mobility in industry and traffic. More information at ai4mobile.org.

Citation

If you use the dataset, please cite it as:

```
@article{hernangomez2022berlin,
    title = {Berlin {{V2X}}: {{A Machine Learning Dataset}} from {{Multiple
Vehicles}} and {{Radio Access Technologies}}},
    shorttitle = {Berlin {{V2X}}},
    author = {Hernang{\'o}mez, Rodrigo and Geuer, Philipp and Palaios, Alex
andros and Sch{\"a}ufele, Daniel and Watermann, Cara and {Taleb-Bouhemadi},
Khawla and Parvini, Mohammad and Krause, Anton and Partani, Sanket and Viel
haus, Christian and Kasparick, Martin and K{\"u}lzer, Daniel F. and Burmeis
ter, Friedrich and Sta{\'n}czak, S{\l}awomir and Fettweis, Gerhard and Scho
tten, Hans D. and Fitzek, Frank H. P.},
    year = {2022},
    month = dec,
    number = {arXiv:2212.10343},
    eprint = {2212.10343},
    eprinttype = {arxiv},
    primaryclass = {cs},
    publisher = {{arXiv}},
    doi = {10.48550/arXiv.2212.10343},
    archiveprefix = {arXiv},
    keywords = {Computer Science - Artificial Intelligence,Computer Science
- Machine Learning,Computer Science - Networking and Internet Architecture}
}
```