An Open Source Wireless Communication Database for Radio Access Network

Yanzan Sun[1], Shengyu Gao[1](✉), Jun Yu[1], Yanyu Huang[1], Shunqin Zhang[1], Xiaojing Chen[1], and Ming Gan[2]

[1] Shanghai University, Shanghai 200444, China
{yanzansun,gaolegao,junyu,huangyanyu,shunqing,jodiechen}@shu.edu.cn
[2] Xintu (Wuxi) New Energy Technology Co. LTD., Wuxi 214000, China
gan.m@xjieg.com

Abstract. The research and development of wireless communication technology is inseparable from the support of experimental dataset. This paper first provides an open source multi-dimensional high-precision database of wireless communication signaling data based on the OAI platform, which can build actual communication scenarios to ensure that the dataset in this database is authentic and reliable. Second, we improve the time precision of this dataset to the millisecond level and realize multi-dimensional wireless communication signaling data collection at both of the physical layer and MAC layer, which can provide a more comprehensive and high-precision data foundation for wireless communication research. Finally, our database is applied for supporting two applications, i.e. RAN slice resource prediction and video bit rate adaptive adjustment, respectively. The experiment results show that our proposed database performs better for supporting the RAN slice resource prediction and video bit rate adaptive adjustment compared with other existing databases.

Keywords: Wireless communication · Database · OAI platform

1 Introduction

Since the advent of the first mobile communication system in the 1980s, wireless communication technology has experienced considerable development in the past few decades. Until the freezing of the 3GPP R16 standard, mobile communication technology has realized the technical iteration from the first generation to the fifth generation, and has now formed a wide range of commercial applications based on three typical application scenarios, i.e. enhanced mobile broadband (eMBB), ultra-reliable low latency communications (uRLLC), and massive machine-type communications (mMTC) [1]. The evolution of wireless communication technology from CDMA to LTE, and then to 5G NR requires a lot of technological improvements and innovations, which cannot be separated from the support of a large number of experimental datasets.

N.-S. Vo and H.-A. Tran (Eds.): INISCOM 2023, LNICST 531, pp. 63–74, 2023.
https://doi.org/10.1007/978-3-031-47359-3_5

In recent years, different types of datasets have been collected for the research of wireless communication technology. Particularly, in [2], a large amount of on-site wireless communication data was collected on the actual HRS (high-speed railway) network to evaluate the impact of frequent handover on TCP protocol deterioration, which was then solved by the proposed congestion control algorithm. Based on time-critical warning information collected from buses, such as 3D spatial information and wireless communication signal data, the authors in [3] studied the expected performance of connected vehicles in LTE networks and estimated the potential benefits of simultaneous multiple access transmission over multiple operators. A large-scale wireless communication dataset at the city level is studied in [4], where the authors measured and made public a rich open multi-source dataset in two geographic regions, namely the city of Milan and the province of Trentino, covering telecommunication, weather, news, social networking, and electricity data, providing an ideal testbed for methodologies and approaches aimed at tackling a wide range of problems such as energy consumption, urban structures and interactions. For both of 5G SA and NSA campus network, the datasets were collected from the campus network to study the packet delays and the losses of one-way transmission in [5], which informed the further development and refinement of 5G SA and NSA campus networks for industrial use cases.

Although the above research are promoting technological progress, there are still some deficiencies in the experimental dataset. First, most of the experimental research is based on the simulation data, without using the real collected data. Second, some studies only use some relevant and incomplete data as the basis for drawing research conclusions. Due to the incompleteness of experimental datasets, the accuracy of research conclusions cannot be guaranteed. Third, the granularities of most of the current datasets are not accurate enough, which inevitably leads to the neglect of features of short time scales. Last, the datasets of many experimental studies are not public due to data protection, so that it is difficult for researchers in the same direction to continue their research on the basis of their predecessors.

In view of the above issues, our work mainly focus on the following aspects:

- *Real Dataset Construction.* Different from the simulation dataset, we construct a real dataset by capturing the real signaling data in actual wireless communication scenarios, which are built by using the OAI (OpenAirInterface) platform. Then the dataset built in this way can be more authentic and reliable.
- *High-precision and Multi-dimensional Dataset Collection.* In order to enable wireless communication research more accurate and comprehensive, we improve the accuracy of the dataset to the millisecond level; meanwhile, we capture multi-dimensional wireless communication signaling dataset in both of the physical layer and the MAC layer. What's more, we open source our high-precision and multi-dimensional database to other wireless communications research peers in [6].

2 Preliminaries

To build a light-weight 5G service delivery platform across reusable software components, OAI provides a system called Mosaic5G. In radio access network, Mosaic5G is mainly composed of OAI-RAN and OAI-CN, and is equipped with network interface tools such as FlexRAN to capture, analyze and control network data in the communication system. In the following, we will describe OAI-RAN, OAI-CN and FlexRAN in detail.

Firstly, OAI-RAN is mainly composed of Openair1, Openair2, Openair3 and Common. Among them, Openair1 corresponds to 3GPP38.321 protocol and consists of the physical layer and physical layer RF simulation module to realize functions such as physical layer coding/decoding, OFDM baseband modulation and demodulation, Fourier transform and inverse transform, etc. Openair2 consists of the entire wireless communication RAN Protocol stacks of layer 2 and layer 3, such as MAC, RLC, PDCP, SDAP and RRC layers, and implements wireless communication data transmission functions including physical layer frame scheduling, radio link transmission modes definition, compression and decompression of IP header, QoS (Quality of Service) to DRB (Data Ratio Bearer) mapping, radio resource management and control, etc. Openair3 includes S1AP and GTP-U modules, and connects the communication between base station and core network in both of user plane and data plane.

Secondly, OAI-CN is for the realization of the 3GPP specifications for the 5G Core Network, and contains the following seven modules: AMF (Access and Mobility Management Function), AUSF (Authentication Server Management Function), NRF (Network Repository Function), SMF (Session Management Function), UDM (Unified Data Management), UDR (Unified Data Repository) and UPF (User Plane Function). Due to the complexity of the 5G in SA mode, the functions and modules of OAI-CN are not perfect and still being supplemented according to 3GPP standards.

Thirdly, FlexRAN platform consists of two main components, i.e. the FlexRAN service and control plane, and the FlexRAN application plane. The FlexRAN service and control plane consists of real-time controllers (RTCs) connecting to multiple underlying RAN runtimes. The separation of control and data planes is achieved through the RAN runtime environment. RAN control applications can be developed on top of the RAN runtime and RTC SDK, and allow monitoring, control and coordination of the state of the RAN infrastructure.

The development and maturity of OAI have driven the research and application in various wireless communication-related fields in both of academia and industry. For example, network slicing [7,8], internet of vehicles [9,10], non-orthogonal multiple access [11], multi-access edge computing [12,13] and other research hotspots in the field of communication are studied based on the open OAI platform, respectively.

Based on the analysis above, quite a few studies have been made based on the deployment of actual OAI systems. However, the wireless communication database research based on OAI is still lacking. Considering that wireless communication database can provide data support for more efficient research, this paper will conduct research on wireless communication dataset acquirement and construction based on the OAI system according to 3GPP series protocol standards.

3 OFDMA Based Wireless Communication Database

OAI RAN has designed a matching RAN communication architecture for different networking modes. In NSA (Non-Standalone) mode, eNB and gNB are connected through X2-C interface; eNB and EPC are connected through S1-U and S1-MME; gNB and EPC are connected through S1-U. In SA (Standalone) mode, the control plane and user plane information of gNB are connected to AMF/UPF via NG-C/U. A schematic diagram is shown in Fig. 1.

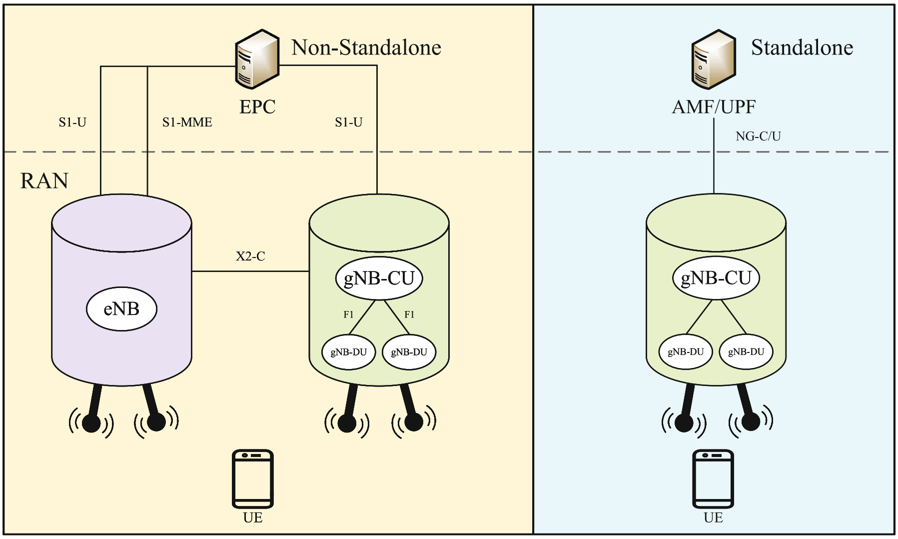

Fig. 1. RAN architecture of OAI system under the dual architecture of NSA and SA.

3.1 Software and Hardware Configuration

According to the existing version of OAI system, considering the stability and performance differences brought about by different construction methods of the OAI system, we built two sets of actual wireless communication scenarios. The first is Separate Mode, where the base station and the core network are deployed on two independent PCs, as shown in Fig. 2; the second one is All In One Mode, where the base station and the core network are deployed on the same PC. In terms of hardware, we use Intel i5 Computer as the platform for the base station and core network, use Huawei Nexus 6P for UE equipment, and adopt Ettus USRP B210 for the RF front end, respectively. AS for software, the host system equipped with the base station and the core network adopts Ubuntu16.04, and the UE system is android8.0. The details of hardware and software arrangement are summarized in Table 1.

Table 1. Details of hardware and software arrangement for wireless communication environment based on OpenAirInterface.

Component	Hardware	Software
UEs	Huawei Nexus 6P	Android 8.0
eNodeB	Intel i5 Computer with low-latency kernel 3.19	OpenAirInterface (OAI) [14]
RF Front-End	Ettus USRP B210	N/A
EPC	Intel i5 Computer with low-latency kernel 3.19	openair-cn [15]

In addition, to obtain the communication dataset under different communication scenarios, we set different service modes for different UEs, such as web browsing, SMS communication, live video with different bit rates, etc, to distinguish the service type and the load differences of UEs.

3.2 Data Collection

We will collect multi-dimensional wireless communication signaling data at both of the physical layer and MAC layer for the dataset construction. Based on the deployment of the above communication scenarios, we will focus on the data transmission of various types on the RAN side. On the one hand, since FlexRAN can directly capture the transmission data of the OAI MAC layer, we are able to capture the transmission data between the UE and the eNB during the actual communication through FlexRAN, including PDU, SUD, etc. On the other hand, by modifying the source code of the OAI MAC layer, we can further capture the physical layer parameters allocated to the UE on the eNB side, such as CQI, RBS, MCS, etc. Parts of the dataset contents are depicted in Fig. 3.

4 Database Analysis

In this section, we will compare our database with several existing wireless communication datasets in terms of time granularity, data integrity, and data authenticity, respectively.

4.1 Time Granularity

In terms of time granularity, we compare our dataset with the public Cellular dataset from Moving Buses, Telecom Italia Big Data Challenge dataset, Link Quality Estimation Data for FlockLab dataset, etc.

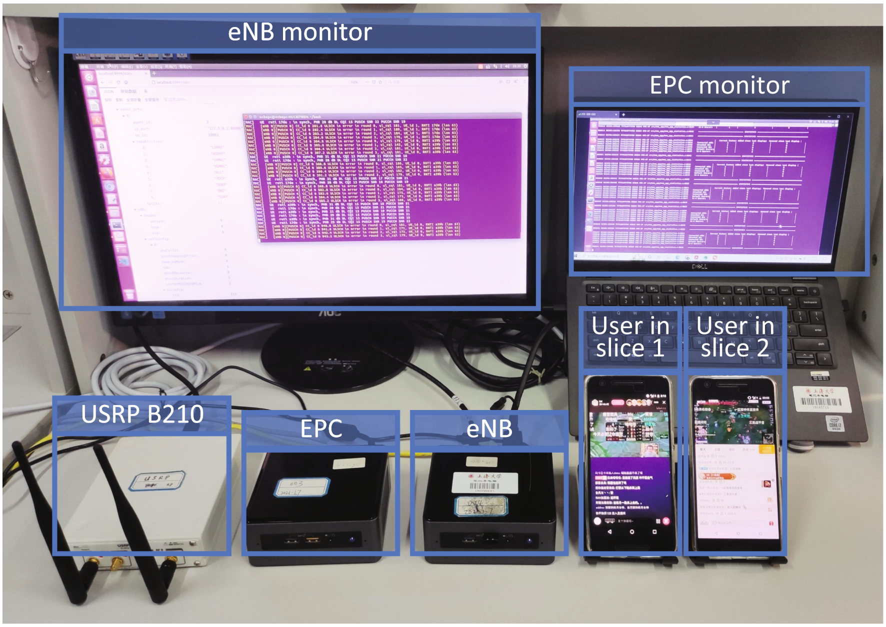

Fig. 2. Actual wireless communication system based on OAI in Separate Mode.

TIME	FRAME	SUBFRAME	UEID	CCID	HARQ	ROUND	RBS	CQI	MCS	RRC	TBS
1630055921267	258	2	0	0	6	8	3	15	28	4	7
1630055921348	266	3	1	0	7	8	3	15	28	4	7
1630055921387	270	2	0	0	6	8	3	15	28	4	7
1630055921468	278	3	1	0	7	8	3	15	28	4	7
1630055921507	282	2	0	0	6	8	3	15	28	4	7
1630055921588	290	3	1	0	7	8	3	15	28	4	7
1630055921627	294	2	0	0	6	8	3	15	28	4	7
1630055921708	302	3	1	0	7	8	3	15	28	4	7
1630055921747	306	2	0	0	6	8	3	15	28	4	7
1630055921828	314	3	1	0	7	8	3	15	28	4	7
1630055921844	315	9	1	0	7	8	3	15	28	4	7
1630055921867	318	2	0	0	6	8	3	15	28	4	7
1630055921874	318	9	1	0	5	8	3	15	28	4	7
1630055921906	322	1	1	0	5	8	3	15	28	4	161

Fig. 3. Part of the dataset content collected by the OAI system.

As shown in Table 2, our dataset can output wireless communication data per millisecond and then reach millisecond-level precision, embodying the advantage of features on the smallest time scale. This means that our database can provide more different time scales data for supporting wireless communication research.

4.2 Data Integrity

In terms of dataset integrity, compared with the public Cellular Received Signal Strength Indicator, LTE Frequency Hopping dataset, etc., our dataset has more comprehensive data types, including Time, UEID, Harq, Hard Round, CQI, MCS, RBS, TBS, PDU, SDU, etc., covering both of the PYH layer and the MAC layer. Furthermore, our underlying dataset also retains the features of the raw data more comprehensively, and can provide input dimension guarantee of multi-dimensional feature input for wireless communication research.

Table 2. Comparison of different wireless communication signaling datasets in terms of time granularity.

dataset	Data Type	Time Resolution
Cellular Received Signal Strength Indicator [2]	Time, Network(2G, 3G, 4G), RSRP, RSS, RSCP, RSSI	30 min
Celluar dataset from Moving Buses [3]	GPS, RSSI, RSRP, Frequency, RSRQ, Band, Protocol, Operator	200 ms
LTE Frequency Hopping dataset [16]	Band, MCS, rblength, UEs, hopinfo	40 ms
LTE KPI [17]	eNB, LCID, RSRP, RSRQ, RSSNR, Carrier Frequency, HandoverDelay, CQI, UE Speed, UE Direction,	1 ms
Link Quality Estimation Data for FlockLab [18]	date_time, rf_channel, tx_power, payload, host_id, rand_seed	2 h
Telecom Italia Big Data Challenge dataset [4]	SMS-in activity, Call-in activity, Internet traffic activity	10 min
Ours	Time, UEID, Harq, CQI, RSRP, RSRQ, Hard Round, RBS, TBS, PDU, SDU, MCS_DL, MCS_UL	1 ms

5 Application

In this section, we use two applications to verify the effectiveness of our proposed wireless communication database.

5.1 Database for Supporting RAN Slicing

We build a RAN slice wireless communication platform, where wireless communication signaling dataset is constructed in single base station, multi-UE, and multi-slice scenarios. Based on this, prediction-based resource slicing is investigated under the premise of ensuring SLA.

The communication scenario is set up with one mini PC building the EPC and one mini PC building the eNB. An USRP B210 is connected to the eNB as a radio frequency transceiver. The terminals are two Huawei Nexus 6Ps, which are connected to the base station for video and web browsing service transmission, respectively. Through FlexRAN, two UEs are assigned to two different slices and run their respective services, where the signaling data are collected.

We set the collected CQI, TBS, MCS as the three input dimension of the CNN-LSTM neural network, to train and predict the number of PRB resources requested by different slices at the next moment. Then slice resources can be allocated in advance to reduce the service SLA level drop caused by inappropriate and untimely allocation of slice resources. Compared with the limitations of existing in dataset integrity and time granularity, we use different input feature dimensions and different time granularity as the independent variables of the experiment to study the variation of SLA satisfaction rate caused by dataset differences. Specifically, we take TBS as the one-dimension input, TBS, MCS as the two-dimension input, TBS, MCS, CQI as the three-dimension input. The time granularity is divided into 100 ms, 500 ms and 1000 ms, respectively.

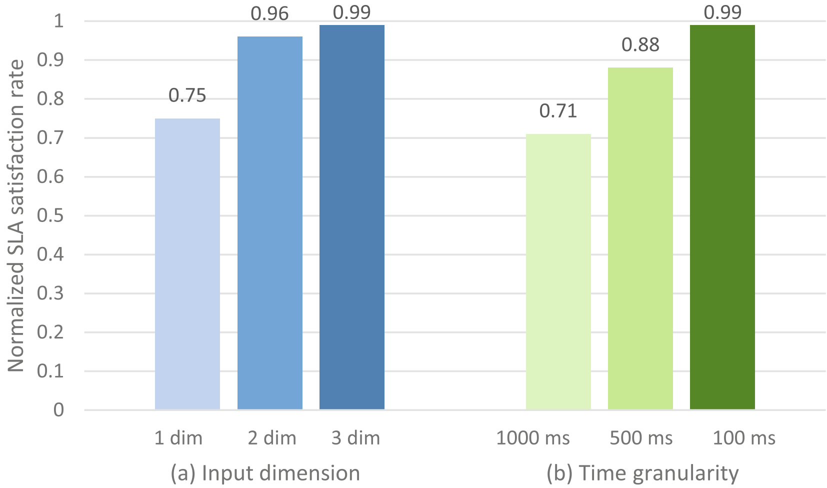

Fig. 4. Comparison of SLA satisfaction rate with different input dimensions and time granularities.

The SLA satisfaction rates with different input dimensions and time granularities are shown in Fig. 4. We can see that the three-dimension input has the highest SLA satisfaction rate and improves the SLA satisfaction rate by 24% over that of one-dimension input. This is due to the fact that a richer input dataset allows the network to learn more complete data characteristics, thereby providing more reasonable decisions for resource allocation after prediction to meet the SLA requirements. On the basis of the three-dimension input, we further compare the experimental results with different time granularities from 1000 ms to 100 ms. Compared with 1000 ms, the time granularity of 100 ms improves the SLA satisfaction rate by 28%. This is because a smaller time granularity allows the network to learn data characteristics on a smaller time scale.

5.2 Database for Supporting Video Bit Rate Adaptive Adjustment

In this subsection, we further verify the effectiveness of our proposed database for supporting video bit rate adaptive adjustment in wireless transmission. We first build a wireless air interface bandwidth prediction and optimization system through FlexRAN, where the real-time parameters data are collected by the interface function in the OAI platform from the PHY layer, the MAC layer, the PDCP layer and other layers of the OAI base station during operation in every 100 ms. In order to keep the data collection time granularity consistent with the bit rate adjustment period, we preprocessed the data, such as mean filtering, feature selection, normalization operations, etc. Second, we input the collected data into LSTM based prediction algorithm to predict the maximum available air interface bandwidth on the user side. Finally, we combine the video quality information returned by the user-side video player, such as the client video buffer, and use the DQN algorithm to optimize the video bit rate adjustment.

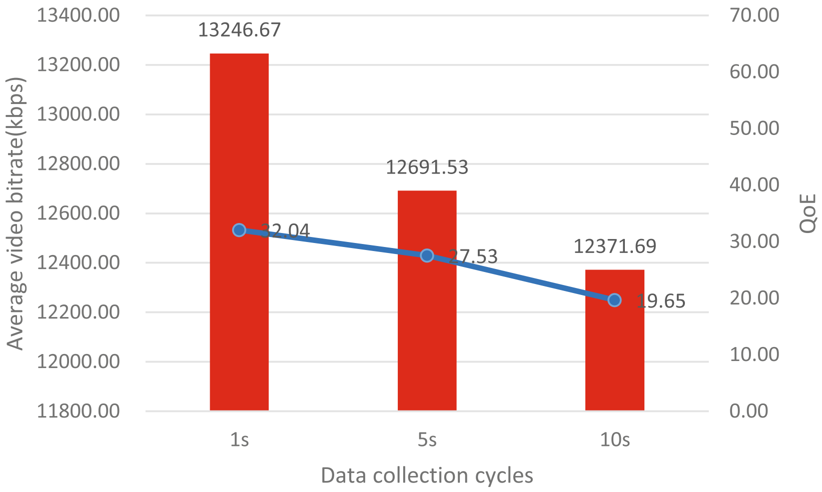

Fig. 5. Performance comparison of different data collection cycles.

The average video bit rate and QoE (Quality of Experience) versus the data collection cycles are shown in Fig. 5. We can see that the performances of the smallest data acquisition period are the best in both of the average video bit rate and QoE. This is because a smaller data acquisition period can provide conditions for a smaller code rate adjustment period. The more frequent video

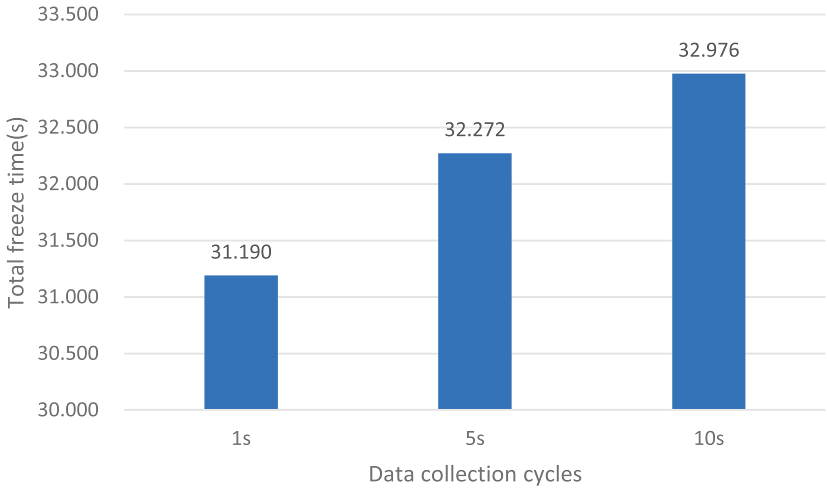

Fig. 6. Comparison of total video freeze time and different data collection cycles.

bit rate switching is also more compatible with the dynamic time-varying channel state variation, so that it can be adjusted to a video bit rate suitable for users more quickly and provide user better QoE. Compared with the 10 s data collection period, the average video bit rate/QoE performance under the 1 s data collection period is 1.07/1.63 times than that of the 10 s data collection period, respectively.

The total video freeze time versus data collection cycles is shown in Fig. 6. The total freeze time corresponding to the 1 s data collection period is 1.786 s less than the total freeze time corresponding to the 10 s data collection period. This is because a larger data collection period corresponds to a larger bit rate adjustment period, which makes The video bit rate fails to respond to the video freeze caused by channel changes in a timely manner.

6 Conclusion

In this paper, we build a real wireless communication database based on the OAI platform. By modifying the OAI source code, a millisecond-level wireless communication signaling dataset can be collected. Furthermore, we built a multi-dimensional high-precision wireless open source database including multi-dimensional wireless communication signaling data in both of the physical layer and the MAC layer. Simulation results of RAN slice resource prediction and video bit rate adaptive adjustment based on our proposed database show that our database provides a good support for wireless communication research.

Acknowledgment. This work was supported by the Innovation Program of Shanghai Municipal Science and Technology Commission under Grant 22511100604 and

20JC1416400, the National Key Research and Development Program of China under Grants 2022YFB2902005, 2022YFB2902304 and 2022YFB2902002, the National Natural Science Foundation of China (NSFC) under Grants 62071284 and 61904101, Key-Area Research and Development Program of Guangdong Province under Grants 2020B0101130012.

References

1. 3GPP: TR 21.916; Services and System Aspects; Release 16 Description; Summary of Rel-16 Work Items (Release 16) (2022)
2. Cui, L., Yuan, Z., Ming, Z., Yang, S.: Improving the congestion control performance for mobile networks in high-speed railway via deep reinforcement learning. IEEE Trans. Veh. Technol. **69**(6), 5864–5875 (2020)
3. Abdesslem, F.B., Abrahamsson, H., Ahlgren, B.: Measuring mobile network multi-access for time-critical C-its applications. In: 2018 Network Traffic Measurement and Analysis Conference (TMA), pp. 1–8 (2018). https://doi.org/10.23919/TMA.2018.8506551
4. Barlacchi, G., et al.: A multi-source dataset of urban life in the city of Milan and the province of Trentino. Sci. Data **2**(1), 1–15 (2015)
5. Rischke, J., Sossalla, P., Itting, S., Fitzek, F.H.P., Reisslein, M.: 5G campus networks: a first measurement study. IEEE Access **9**, 121786–121803 (2021). https://doi.org/10.1109/ACCESS.2021.3108423
6. LabGreat: OAI dataset. https://github.com/LabGreatOfficial/OAI. Accessed 17 Mar
7. Liu, Q., Han, T., Moges, E.: EdgeSlice: slicing wireless edge computing network with decentralized deep reinforcement learning. In: 2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS), pp. 234–244 (2020). https://doi.org/10.1109/ICDCS47774.2020.00028
8. Schmidt, R., Chang, C.Y., Nikaein, N.: FlexVRAN: a flexible controller for virtualized ran over heterogeneous deployments. In: ICC 2019–2019 IEEE International Conference on Communications (ICC), pp. 1–7 (2019). https://doi.org/10.1109/ICC.2019.8761222
9. Manco, J., Baños, G.G., Härri, J., Sepulcre, M.: Prototyping V2X applications in large-scale scenarios using OpenAirInterface. In: 2020 IEEE Vehicular Networking Conference (VNC), pp. 1–4 (2020). https://doi.org/10.1109/VNC51378.2020.9318327
10. Gill, K.S., et al.: Bumblebee-inspired C-V2X dynamic spectrum access testbed using OpenAirInterface. In: 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), pp. 1–5 (2020). https://doi.org/10.1109/VTC2020-Spring48590.2020.9128871
11. Wei, X., et al.: Software defined radio implementation of a non-orthogonal multiple access system towards 5G. IEEE Access **4**, 9604–9613 (2016)
12. Li, C.Y., et al.: Transparent AAA security design for low-latency MEC-integrated cellular networks. IEEE Trans. Veh. Technol. **69**(3), 3231–3243 (2020)
13. Li, C.Y., et al.: Mobile edge computing platform deployment in 4G LTE networks: a middlebox approach. In: USENIX Workshop on Hot Topics in Edge Computing (HotEdge 2018). USENIX Association, Boston, MA (2018). https://www.usenix.org/conference/hotedge18/presentation/li
14. Openairinterface repository. https://gitlab.eurecom.fr/oai/openairinterface5g. Accessed 10 Oct 2021

15. OpenAir-CN repository. https://gitlab.eurecom.fr/oai/openair-cn. Accessed 17 Oct 2021
16. Cintron, F.: Performance evaluation of LTE device-to-device out-of-coverage communication with frequency hopping resource scheduling (2018-07-23 2018). https://doi.org/10.6028/NIST.IR.8220
17. Mohan, D., Geetha Mary, A.: LTE KPI for data mining and machine learning (2020). https://doi.org/10.17632/czkn9c4wk6.2
18. Jacob, R., Forno, R.D., Trüb, R., Biri, A., Thiele, L.: Dataset: wireless link quality estimation on FlockLab - and beyond. In: Proceedings of the 2nd Workshop on Data Acquisition to Analysis, DATA 2019, pp. 57–60. Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3359427.3361907