A Large-Scale Dataset of 4G, NB-IoT, and 5G Non-Standalone Network Measurements

Konstantinos Kousias*, Mohammad Rajiullah†, Giuseppe Caso†, Usman Ali‡, Ozgu Alay*†, Anna Brunstrom†, Luca De Nardis‡, Marco Neri§, and Maria-Gabriella Di Benedetto‡

*University of Oslo, Norway
†Karlstad University, Sweden
‡Sapienza University of Rome, Italy
§Rohde & Schwarz, Italy

Abstract—**Mobile networks are highly complex systems. Therefore, it is crucial to examine them from an empirical perspective to better understand how network features affect performance, so to suggest additional improvements. To this aim, this paper presents a large-scale dataset of measurements collected over fourth generation (4G) and fifth generation (5G) operational networks, providing Long Term Evolution (LTE), Narrowband Internet of Things (NB-IoT), and 5G New Radio (NR) connectivity. We collected our dataset during seven weeks in Rome, Italy, by performing several tests on the infrastructures of two major mobile network operators (MNOs). The open-sourced dataset has enabled multi-faceted analyses of network deployment, coverage, and end-user performance, and can be further used for designing and testing artificial intelligence (AI) and machine learning (ML) solutions for network optimization.**

I. INTRODUCTION

The ultimate goal of current and future cellular systems is to enable an increasing number of services over the mobile network. In the fifth generation (5G) era, enhanced mobile broadband (eMBB), ultra-reliable low-latency communication (URLLC), and massive machine-type communication (mMTC) require stringent quality of service (QoS) in terms of throughput, latency, and connection reliability.

Considering the systems standardized by the 3rd Generation Partnership Project (3GPP), initial steps of the ongoing evolution of the cellular ecosystem were Release 15 (Rel-15) and Release 13 (Rel-13), where 5G and Narrowband Internet of Things (NB-IoT) systems were defined, with their enhancements addressed in the following releases.

The interworking between 5G, NB-IoT, and pre-existing fourth generation (4G) Long Term Evolution (LTE) and LTE-Advanced (LTE-A) systems enables flexible and efficient support for eMBB, URLLC, and mMTC. Therefore, 5G and NB-IoT deployment on top of 4G networks is rapidly progressing worldwide. This allows for conducting empirical investigations in parallel to theoretical analyses, considering that data-driven research is crucial toward identifying how performance is affected by deployment, configuration, and technology features, and how it can be further improved.

The need for data-driven analyses is also fueled by the currently popular approach of embedding artificial intelligence (AI) and machine learning (ML) in the management and optimization of several network aspects. This approach requires the use of trustable datasets, either generated synthetically or collected over real systems, toward properly designing AI/ML solutions and showcasing their benefit.

Within the above context, this paper presents an open-source dataset collected during a large-scale measurement campaign carried out in Rome, Italy, on the networks of two mobile network operators (MNOs) providing 4G, 5G, and NB-IoT connectivity in low-to-mid frequency bands (0.8–4 GHz) [1].

As detailed later, parts of the dataset were already used to support specific analyses [2]–[4]; other relevant work can be found in [5] for 4G, [6]–[8] for 5G, and [9]–[11] for NB-IoT, where data collection and analyses were carried out on each technology separately, in some cases finalized to propose AI/ML solutions for specific aspects

(e.g., 5G throughput prediction [6]–[8] and NB-IoT access protocol optimization [10]).

With respect to the above work and to the best of our knowledge, our work discloses the first city-wide dataset in a European country that is, as described in the following, *multi-technology and multi-operator* (4G, 5G, NB-IoT networks of two MNOs), *multi-device* (network scanner for passive monitoring and user device for active performance tests), *multi-scenario* (indoor static and outdoor mobile tests), and *multi-service* (eMBB/URLLC-like performance tests). The dataset enables in-depth analyses of coverage and performance across technologies, geographical areas, scenarios, frequency bands, and MNO infrastructures, ultimately allowing the design and test of AI/ML solutions for multiple network aspects.

The article is organized as follows: first, we describe relevant 5G and NB-IoT deployment aspects, followed by an overview of the data collection process. We then provide a description of the dataset, present potential AI/ML use cases for it, and conclude our work.

II. BACKGROUND

This section describes 5G and NB-IoT aspects relevant to the dataset presented in this paper.

As anticipated above, MNOs are deploying 5G and NB-IoT using roll-out options that allow for a smooth integration with their 4G networks.

5G can be deployed in non-standalone (NSA) or standalone (SA) mode [12]. For the initial deployment, the majority of MNOs are adopting the NSA mode (use of the 4G core network (CN)), with SA (use of the 5G CN) expected to prevail in the long-term. In both cases, the 5G radio access network (RAN) is formed by Next Generation Node Bs (gNBs) and corresponding New Radio (NR) cells identified by their Physical Cell IDs (PCIs), i.e., identifiers reusable across the RAN (similar to 4G E-UTRAN Node Bs (eNBs) and PCIs).

Synchronization Signal Block (SSB) beamforming and dual connectivity (DC) are two main 5G features for performance improvement. SSB beamforming uses multiple beams (up to 8 in the 5G mid-band, as per Rel-15) to direct signals from a PCI toward different users, thus increasing spatial diversity. DC allows a user equipment (UE) to simultaneously connect to a 4G and a 5G PCI, thus increasing throughput and reliability.

NB-IoT can be deployed either in guard-band, in-band, or standalone mode [13]. In guard-band, NB-IoT uses a Physical Resource Block (PRB) of 180 kHz in the guard band among different sets of LTE PRBs. In in-band, the NB-IoT PRB is within the LTE band, while in standalone, NB-IoT uses a 200 kHz channel in the Global System for Mobile Communications (GSM) spectrum. After selecting a mode, NB-IoT services are provided by upgrading 4G eNBs and PCIs.

III. DATA COLLECTION

This section outlines the setup, measurements and mobility scenarios adopted to collect our dataset.

A. Setup

The measurement campaign took place in Rome, Italy, during seven weeks between Dec. 2020 and Jan. 2021. For active measurements, we used the Rohde & Schwarz (R&S) Android application *Qualipoc* running on a Samsung S20 5G-capable UE, while for passive measurements, we used the R&S TSMA6 toolkit, a system composed by an Intel PC (Windows) and a spectrum scanner. For network analysis, troubleshooting, visualization, and data exporting, we used the R&S ROMES software. The complete setup is shown in Figure 1, and consisted of an omnidirectional radio frequency (RF) antenna (i), a synchronized Global Positioning System (GPS) antenna (ii), the TSMA6 (iii), and the UE (iv). A tablet (not shown in the figure) was used to connect to the PC embedded in the TSMA6.
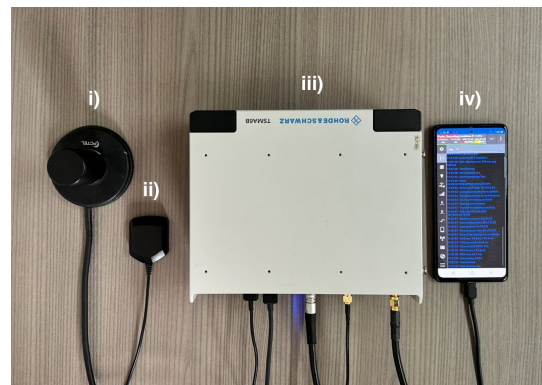

Fig. 1: The measurement setup consisting of i) an RF antenna, ii) a GPS antenna, iii) the R&S TSMA6, and iv) a 5G-capable UE.

B. Measurements

The above setup allowed us to perform both passive network monitoring and active performance tests, as detailed below.

Passive monitoring: the TSMA6 was used to decode downlink control signals from all the cells operating in four LTE Bands (1, 3, 7, and 20), one NB-IoT Band (guard band of LTE Band 20), and one 5G NR Band (n78).

4G and NB-IoT measurements were reported at the PCI level, while 5G measurements were reported at the SSB level, i.e., up to eight samples for 5G PCIs adopting SSB beamforming. The high sensitivity of the scanner allowed to decode rather weak signals from distant cells, which a less sensitive UE may not be able to decode, thus enabling very granular analyses on deployment and coverage.

Active performance tests: We used the UE for executing throughput and delay/reliability performance tests on the 4G and 5G NSA networks of two Italian MNOs, denoted as Op_1 and Op_2. The UE was configured in two different modes. In *5G-disabled* mode, the UE only exposed 4G capabilities so that it could only connect to 4G PCIs; In *5G-enabled* mode, the UE exposed 5G NSA capabilities so that it could connect to 5G and 4G PCIs.

We organized the campaign into *sub-campaigns* carried out during different days and times of day. For each sub-campaign, we repeated the performance tests several times; in the following, we use the term *session* to refer to each repetition. Next, we give an overview of the executed tests.

1) Throughput Test: We used the Speedtest by Ookla jointly with R&S *Qualipoc* to measure the end-to-end downlink (DL) and uplink (UL) throughput in 5G-enabled and 5G-disabled modes. The target server was located in Rome, and each session lasted about 60 seconds. We opted for the Speedtest multi-connection mode, which uses Transmission Control Protocol (TCP) and a proprietary algorithm for determining the number of connections [2].

2) Latency/Reliability Test: To assess latency and reliability, we used the *Qualipoc* interactivity test with the real-time online gaming traffic pattern. Each test included multiple 10-second long sessions, during which various game interactivity phases were replicated, with data rates from 0.1 to 1 Mbps [2]. The test allowed for setting a delay budget on the exchanged User Datagram Protocol (UDP) packets, so that packets not received within the budget were considered lost by the client. We used a 100 ms budget, as specified by 3GPP for this application class in Rel-15 Technical Specification 23.501.

C. Scenarios

In order to assess the influence of mobility and location, we defined three sub-campaign scenarios: *indoor static (IS)*, for data collected at the seventh floor of a residential building, and different offices at the second floor of the Department of Information Engineering, Electronics and Telecommunications (DIET) of Sapienza University of Rome; *outdoor walking (OW)*, for data collected outdoor while walking (3.47 km/h average speed); and *outdoor driving (OD)*, for data collected outdoor while driving a car (18.54 km/h average speed). In both mobility scenarios, the average speed was affected by factors such as traffic lights and road congestion.

IV. DATASET

This section provides a description of our dataset, a summary of its statistical information, and examples of the collected data. Due to space limitations, we group the features into classes, and provide a detailed, formal feature description in [1].

A. Passive dataset

The passive dataset includes three character-separated values (CSV) files (one per technology).

Table I shows the available feature classes, which include spatial and temporal fields, frequency and cell identifiers, and signal strength and quality indicators measured on 4G Reference Signal (RS), NB-IoT RS, and different 5G control signals and channels. We also included campaign and scenario features that can be used to isolate particular sub-campaigns. The same labeling scheme is used in the active dataset so to enable joint passive/active analyses. For 4G and NB-IoT, we engineered additional features, i.e., the line of sight (LoS) distance between the UE and a detected cell and the number of cells associated to each eNB. An estimated position of 5G PCIs can also be inferred by using online databases, e.g., www.lteitaly.it (Accessed on: March 2023), while also considering that MNOs mostly deploy 5G PCIs on top of 4G RAN sites.

Figure 2 shows an example of the dataset, i.e., the highest Secondary Synchronization Signal (SS) RSRP measured across the 5G PCIs of Op_1.

Feature Class	Description
Temporal information	Date [dd.mm.yyyy] and time [HH:MM:SS.sss].
Spatial and mobility information	UE and cell* GPS coordinates [°]; UE speed [km/h].
Mobile network information	Includes Carrier Frequency [MHz], E-UTRA Absolute Radio Frequency Channel Number (EARFCN), Band identifier, and mobile network code (MNC).
Cell site information	Includes PCI, cell ID (CID)*, eNB* and 5G SSB beam~ identifiers.
Radio coverage information	Includes Reference Signal Received Power (RSRP)$\hat{~}$ [dBm], Reference Signal Received Quality (RSRQ)$\hat{~}$ [dB], Signal to Interference and Noise Ratio (SINR)$\hat{~}$ [dB], and received Power$\hat{~}$ [dBm].
Campaign information	Scenario (IS/OW/OD) and measurement sub-campaign identifier.

* Only available for 4G and NB-IoT.
~ Only available for 5G.
$\hat{~}$ The NB-IoT dataset provides such features for two antenna ports, referred to as `Tx1` and `Tx2`, while the 5G dataset covers different types of signals, i.e., SS, Demodulation Reference Signal (DMRS), Physical Broadcast Channel (PBCH), Primary Synchronization Signal (PSS), and SS-PBCH.

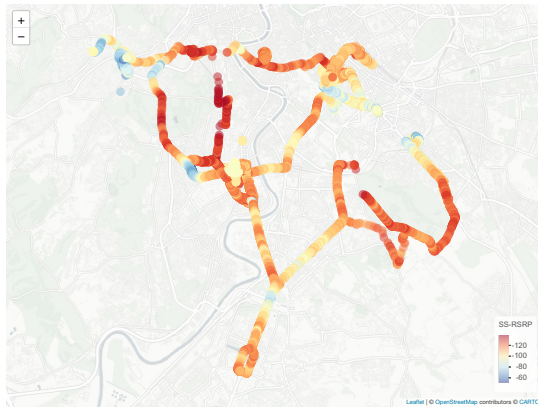TABLE I: Passive dataset feature classes along with a short description.

Fig. 2: Highest SS-RSRP [dBm] measured across the 5G PCIs of Op_1 detected at the locations traversed during *OW* and *OD* sub-campaigns.
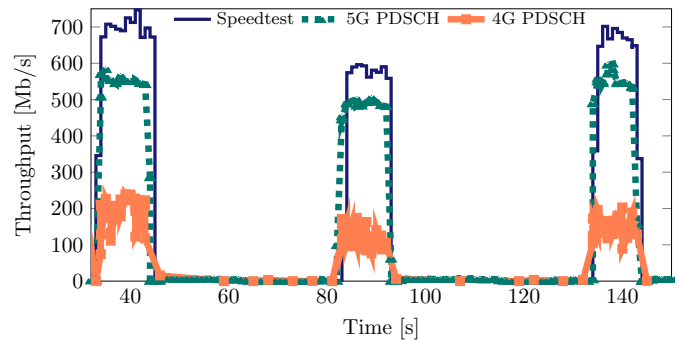

Fig. 3: DL throughput at application layer (using DC), 5G PDSCH, and 4G PDSCH, measured during 3 Speedtest sessions in an *IS* sub-campaign for Op_1.

B. Active dataset

Our active dataset includes two CSV files (one per performance test). Table II lists the feature classes, which can be categorized in two groups: (i) indicators collected by the UE related to radio and physical layers and (ii) QoS and quality of experience (QoE) indicators obtained from running the tests. We also added features related to UE mode, scenario, MNO, and sub-campaign name.

During the campaign, the dataset was updated (i.e., a new row was formed) every time a new value for a feature was collected, with an update time at millisecond-level granularity. Therefore, each row in the files only contains values for the updated features, while ? represents unaltered features.

As an example, Figure 3 shows the Speedtest DL throughput achieved during three 5G-enabled sessions executed in an *IS* campaign for Op_1. The figure shows the use of DC under Op_1 NSA infrastructure, with a high throughput achieved at the application layer via simultaneous use of 5G and 4G PDSCH.

C. Statistics for passive and active dataset

The number of samples in the 4G, 5G, and NB-IoT passive datasets are approximately 528K (243K for Op_1 and 285K for Op_2), 8.14M (1.12M for Op_1 and 7.02M for Op_2), and 281K (133K for Op_1 and 148K for Op_2), respectively. During our measurement campaign, only Op_2 was adopting SSB beamforming, resulting in a higher number of 5G samples collected for Op_2 compared to Op_1.

As regards to the active measurements, we conducted 555 Ookla Speedtest sessions (197 for Op_1 and 358 for Op_2) and 1158 real-time online gaming sessions (657 for Op_1 and 501 for Op_2).

V. AI/ML APPLICATIONS AND USE CASES

We now outline four examples of AI/ML use cases where our dataset can be used, and put them

Feature Class	Description
Temporal information	Date [dd.mm.yyyy] and time [HH:MM:SS.sss].
Spatial information	UE's GPS coordinates [°].
Connection information	Includes PCI of the serving cell(s) and indication if the UE is connected to a 5G and/or a 4G cell.
5G beam information	Number of detected SSB beams of the serving 5G cell and the ID of the serving SSB beam.
Radio coverage information	Includes RSRP [dBm], RSRQ [dB], and SINR [dB] of the serving cell(s).
Resource allocation information	Includes Transport Block Size (TBS) [bit (4G), bytes (5G)], number of resource blocks (RBs) assigned in DL/UL, Modulation and Coding Scheme (MCS), indication of the usage of different modulation schemes (e.g., 256-QAM), indication of the usage of carrier aggregation (CA) for LTE DL (e.g., number of component carriers (CCs)).
Lower-layer throughput information	Includes throughput measured at Physical Downlink Shared Channel (PDSCH) and Physical Uplink Shared Channel (PUSCH) [kb/s].
Higher-layer (application) throughput information*	Includes instantaneous, average, and maximum DL/UL throughput during and at the end (average) of each Speedtest session [kb/s].
Packet statistics⁂	Total number of packets sent, not received in time, during a session.
Packet delay statistics⁂	Round Trip Time (RTT) and packet delay variation (PDV) of the packets sent during a session (median and 10th percentile values) [s].
Interactive performance⁂	Interactivity score (i-score) as a function of RTT, PDV, and packet error rate [%]. For more details, see [2] and https://tinyurl.com/4fzemjff (Accessed on: March 2023).
Campaign information	UE mode (5G-enabled/-disabled), scenario (IS/OW/OD), MNOs, and campaign identifier.

* Available for throughput test.
⁂ Available for latency/reliability test. The dataset contains values measured during and at the end of each session.

TABLE II: Active dataset feature classes along with a short description.

in the context of relevant research work.

A. User/device Positioning

User/device positioning based on cellular networks is receiving increasing attention in research and standardization communities, due to the high expectation toward cellular-enabled location-based services [12]. Focusing on 5G, a recent review on 5G positioning presented in [14] clearly highlights the need for data from real-world 5G networks, as it points out that nearly all proposed techniques are evaluated on simulated data or, at best, on data collected in ad-hoc testbeds. At least seven different studies are identified in [14] where ML positioning techniques based on Neural Networks (NNs), k-Nearest Neighbour (kNN), Deep Neural Networks (DNNs), Support Vector Machines (SVMs), and Gaussian Processes (GPs) are proposed, and only tested on simulated 5G coverage data (e.g., RSRP). Positioning is thus a key use case for our dataset, that can enable a realistic assessment of the accuracy of the above techniques by leveraging over 8 Million data entries, as highlighted in the previous section.

The availability of data on three technologies (NB-IoT, LTE, 5G NR) collected at the same time in the same locations is an additional unique feature of our dataset with respect to positioning; on the one hand, it allows to reliably assess and compare the positioning accuracy provided by different technolo-
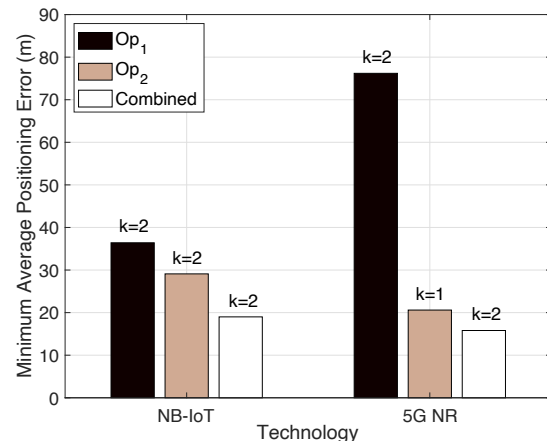

Fig. 4: Minimum Average Positioning Error obtained for NB-IoT and 5G by using the WkNN fingerprinting technique introduced in [11]. The figure also reports, for each combination of technology and operator (Op_1, Op_2, and both combined), the value of k minimizing the average positioning error.

gies and, on the other hand, it enables the definition of multi-technology ML models for positioning.

As an example, Figure 4 depicts the positioning accuracy achieved by adopting weighted kNN (WkNN) fingerprinting on NB-IoT and 5G NR coverage data (i.e., SINR in this instance) collected during the same sub-campaigns. The accuracy is evaluated in terms of the Minimum Average Positioning Error, i.e., the minimum positioning error as

a function of k, for each combination of technology and operator. The results for NB-IoT are a subset of those presented in [11], where a full description of used measurements, proposed WkNN fingerprinting strategies, and adopted data preprocessing (e.g., data for each PCI was spatially smoothed using the 40λ-rule to mitigate fast fading effects) are available. Figure 4 shows that, for Op_1, NB-IoT leads to higher accuracy than 5G due to the higher density of PCIs detected at each location, guaranteed by the larger NB-IoT radio coverage. For Op_2, however, the use of SSB beamforming leads to multiple signals associated to the same PCI being detected at each location, considerably increasing the data density; as a result, 5G outperforms NB-IoT. For both technologies, the combination of data collected for the two MNOs improves the positioning accuracy, in agreement with the conclusions drawn in [11].[1]

B. Propagation Channel Modeling

A well-recognized AI/ML use case for communication systems is the modeling of wireless propagation channel characteristics, toward overcoming the accuracy limitation of simplistic empirical models and the complexity of deterministic models. A relevant example is given in [15], where a channel modeling framework based on NNs was proposed for predicting propagation properties, e.g., received power and root mean square delay spread, using as input the information on transmitter/receiver positions and carrier frequency. The framework showed higher accuracy when applied to real data rather than to synthetic data, thus highlighting that complex channel characteristics can be learned via ML, and open-source datasets are key for further investigating this approach.

The dataset open-sourced in this paper, and the smaller companion dataset disclosed in [4], can provide a key contribution in this context, thanks to the large quantity of heterogeneous data, e.g., in terms of scenarios, frequencies, and technologies, that can be used for model training, validation, and refinement. To give an example referring to Table I, features in *Spatial and mobility information* class (e.g., UE and cell coordinates), *Mobile network information* class (e.g., carrier frequency), and *Cell site information* class (e.g., cell/beam identifiers)

[1]The code for the analysis presented in this Section is available at https://codeocean.com/capsule/1679985/tree/

can be used as input for deriving a ML model for the features in the *Radio coverage information* class (e.g., RSRP). The accuracy and complexity of such a model can be then compared against traditional models leveraging same or different sets of features.

C. QoS Prediction

The predictive modeling of QoS parameters, e.g., throughput and latency, constitutes another AI/ML use case for our dataset. Good estimates of such indicators are essential for operations such as traffic management and network optimization (e.g., resource allocation and user scheduling). In the literature, [6] presented a ML-based framework that leverages gradient boosted decision trees (GBDT) and long short-term memory (LSTM) models for 5G throughput prediction by considering location, mobility, and cell features, while [7] performed a study of client-based throughput prediction for 5G NSA vehicle-to-cloud communications.

In this context, the rich variety of features in our dataset can further allow the application of AI/ML for QoS prediction, toward disclosing attributes and configurations of 5G NSA networks that should be considered for accurate predictions. As an example, features in the *Radio coverage information* class (e.g., RSRP, RSRQ, SINR) can be used as input variables to a ML model for predicting features in the *Higher-layer throughput information* class (e.g., application DL/UL throughput) or *Interactive performance* class (e.g., RTT).

D. Handover Prediction

The analysis and modeling of vertical and horizontal handovers (HOs) is another important aspect to further investigate in 5G networks, considering that the complex nature of HO events results in significant implications on user performance. We preliminary highlighted this aspect in [2], where we leveraged part of our dataset to showcase the impact of HOs on coverage and latency. Another relevant work can be found in [8], where a system that leverages mobility data from 5G cells for HO forecasting was proposed, toward improving 6K panoramic video on demand and real-time volumetric video streaming applications.

The *multi-device* nature of our dataset, which allowed for the simultaneous collection of measurements from multiple 4G and 5G cells, can enable

the design of ML algorithms for HO prediction, to be leveraged for promptly preparing HO-related network resources, e.g., toward HO signaling optimization. For example, active features such as the PCI and connectivity mode in the *Connection information* feature class can be used as input to a time-series model for predicting when vertical/horizontal HOs will happen. Additional passive data, e.g., in the *Radio coverage information* features class, can also be used to enhance the prediction accuracy.

VI. Conclusions

This paper presents a large-scale dataset of 4G, 5G, and NB-IoT measurements collected over the network infrastructures of two large MNOs during a period of seven weeks in Rome, Italy. The dataset offers a variety of network features collected both by leveraging active but also passive measurements, and has been used for the study and analysis of aspects such as radio coverage, deployment, end-user performance, outdoor user/device positioning, and HO analysis. We open-source the dataset to allow for further exploration and analysis and for use with AI and ML use cases.

Acknowledgments

Biographies

Konstantinos Kousias is a Postdoctoral Fellow at University of Oslo. His research focuses on the empirical modeling and evaluation of mobile and IoT systems using data analytics and AI.

Mohammad Rajiullah is a Senior Lecturer at Karlstad University. His research interests are in the areas of low latency networking, web performance, mobile networks, and IoT.

Giuseppe Caso is a Senior Lecturer at Karlstad University. His research interests include cognitive communications, mobile cellular systems and IoT technologies, and location-based services.

Usman Ali is a Ph.D. student at Sapienza University of Rome. His research interests include rate-splitting multiple access and MIMO.

Özgü Alay is a Full Professor at University of Oslo. Her research interests include mobile networks, low latency networking, multipath protocols, and multimedia transmission over wireless networks.

Anna Brunstrom is a Full Professor and Research Manager for the Distributed Systems and Communications Research Group at Karlstad University. Her research interests include Internet architectures and protocols, low latency communication, multipath communication and performance evaluation of mobile systems.

Luca De Nardis is an Associate Professor at the DIET Department at Sapienza University of Rome. His research interests focus on indoor positioning, UWB and cognitive communications, and routing.

Marco Neri joined R&S in 2017 as Application Engineer for Mobile Network Testing. His focus is on 5G and IoT testing worldwide.

Maria-Gabriella Di Benedetto is a Full Professor at the DIET Department at Sapienza University of Rome. Her research interests are focused on wireless communications, speech, and signal processing.

References

[1] K. Kousias *et al.*, "A Large-Scale Dataset of 4G, NB-IoT, and 5G Non-Standalone Network Measurements," 2022. [Online]. Available: https://dx.doi.org/10.21227/7a8s-nt68

[2] K. Kousias *et al.*, "Implications of Handover Events in Commercial 5G Non-Standalone Deployments in Rome," in *Proceedings of the ACM SIGCOMM Workshop on 5G and Beyond Network Measurements, Modeling, and Use Cases*, 2022, pp. 22–27.

[3] K. Kousias *et al.*, "Coverage and Performance Analysis of 5G Non-Standalone Deployments," in *Proceedings of the ACM MobiCom Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization*, 2022, pp. 1–8.

[4] U. Ali *et al.*, "Large-Scale Dataset for the Analysis of Outdoor-to-Indoor Propagation for 5G Mid-Band Operational Networks," *Data*, vol. 7, no. 3, p. 34, 2022.

[5] V. Raida *et al.*, "Real World Performance of LTE Downlink in a Static Dense Urban Scenario - An Open Dataset," in *Proceedings of the IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[6] A. Narayanan *et al.*, "Lumos5G: Mapping and Predicting Commercial mmwave 5G Throughput," in *Proceedings of the ACM Internet Measurement Conference*, 2020, pp. 176–193.

[7] B. Sliwa *et al.*, "Machine Learning-Enabled Data Rate Prediction for 5G NSA Vehicle-to-Cloud Communications," in *Proceedings of the IEEE 5G World Forum*. IEEE, 2021, pp. 299–304.

[8] A. Hassan *et al.*, "Vivisecting Mobility Management in 5G Cellular Networks," in *Proceedings of the ACM SIGCOMM 2022 Conference*, 2022, pp. 86–100.

[9] C.-W. Yau *et al.*, "NB-IoT Coverage and Sensor Node Connectivity in Dense Urban Environments: An Empirical Study," *ACM Transactions on Sensor Networks (TOSN)*, 2022.

[10] G. Caso *et al.*, "NB-IoT Random Access: Data-driven Analysis and ML-based Enhancements," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11 384–11 399, 2021.

[11] L. De Nardis *et al.*, "Positioning by Multicell Fingerprinting in Urban NB-IoT Networks," *Sensors*, vol. 23, no. 9, 2023.

[12] E. Dahlman *et al.*, *5G NR: The Next Generation Wireless Access Technology*. Academic Press, 2020.

[13] O. Liberg *et al.*, *Cellular Internet of Things: From Massive Deployments to Critical 5G Applications, 2nd Edition*. Academic Press, 2020.

[14] F. Mogyorósi *et al.*, "Positioning in 5G and 6G Networks – A Survey," *Sensors*, vol. 22, no. 13, 2022.

[15] J. Huang *et al.*, "A Big Data Enabled Channel Model for 5G Wireless Communication Systems," *IEEE Transactions on Big Data*, vol. 6, no. 2, pp. 211–222, 2018.