Poster: LoRa Mobility and Coverage Dataset (LoRaMC)

Lorenzo Frangella
Stefano Milani
frangella.1899674@uniroma1.it
stefano.milani@uniroma1.it
University of Roma "La Sapienza"
Rome, Italy

Domenico Garlisi
University of Palermo
CNIT, National Inter-University Consortium for Telecommunications
Palermo, Italy
domenico.garlisi@unipa.it

Ioannis Chatzigiannakis
University of Roma "La Sapienza"
CNIT, National Inter-University Consortium for Telecommunications
Rome, Italy
ichatz@diag.uniroma1.it

ABSTRACT

The rapid growth of the Internet of Things (IoT) and the emergence of Low Power Wide Area Network (LPWAN) technologies, such as LoRaWAN, have revolutionized how applications and services can leverage sensor and actuator devices. The truthful evaluation of applications and services relying on large-scale IoT deployments relies on the use of simulation tools in combination with datasets developed from real-world traces. In this work, we introduce a new dataset suitable for evaluating scenarios in urban environments that may involve mobile end-devices. The dataset includes an enriched taxi mobility data generated from 3300 mobile devices moving within a 200 square kilometer area that corresponds to the metropolitan city of Rome. The resulting dataset indicates for each device, it is position, which GWs received the frames, and with what radio characteristics. We believe that such a dataset can be used to highlight the effectiveness, scalability, and robustness of the LPWAN based system in the city-scale scenarios as well as help train and evaluate machine learning and artificial intelligence based methods that are embedded in network operation and resource optimization layers.

CCS CONCEPTS

• Networks → Network architectures; **• Computing methodologies → Distributed computing methodologies**; **• Computer systems organization** → *Real-time systems.*

KEYWORDS

IoT, LoRaWAN, Edge-to-Cloud Continuum

1 INTRODUCTION

The Internet of Things (IoT) has been a game changer in the way we operate, live, commute, and conduct business thanks to its capacity to monitor in real-time a broad range of environmental parameters in indoor, outdoor, industrial, urban, as well as rural areas. Within this diverse landscape of massive IoT use cases, emerging Low-Power Wide-Area Network (LPWAN) technologies introduce a uniform approach for creating global networks of devices for power-efficient and cost-effective integration of IoT data streams with cutting-edge web services across numerous specialized sectors. Today, LPWAN represents one of the most rapidly deployed long-range communication architectures to facilitate large-scale connectivity among IoT devices, potentially reaching a staggering 22 billion connections by 2025 [1].

In this work, we focus on Long-Range Wide-Area Network (LoRaWAN) as one of the most adopted LPWAN technologies. LoRaWAN introduces a centralised, monolithic, architecture where network gateways (GWs) bridge traffic arriving from the IoT devices with the network backbone by directly forwarding all the IoT traffic to central cloud services [2]. Following a licence-free ad-hoc deployment model, it provides an ideal solution for connecting a wide range of IoT devices with minimal infrastructure requirements. At the same time, the standard producer/consumer model adopted by LoRaWAN allows the straight-forward deployment of cloud-based analytics services to process the large amounts of IoT generated data streams.

The research community has traditionally relied on three main approaches to carry out the experimental-driven evaluation, including in the IoT landscape: *physical testbeds*, *emulation* and *simulation*. Today the use of physical testbeds, like for example IoT-Lab [1] [3] offer the possibility to evaluate the performance of the firmware when operated at a single device. Towards moving to large-scale evaluation of the system involving hundreds of devices, simulation remains the only tool available. The use of simulation involves the development of synthetic scenarios or the use of datasets that are based on traces collected from real-world deployments. The use of simulation is also important in view of data-driven analysis that are required within the context of embedding machine learning and artificial intelligence at network operation and resource optimization level.

For LPWAN technologies, and in particular LoRaWAN, the availability of public datasets is still very limited and those available are often outdated: many of the measurement campaigns that were made available in recent years were performed on 20 or 40 MHz frequency bands, and do not contain sufficient *domain diversity* so that they can be reliably used for simulating at large-scale LPWAN technologies.

In this work, we present LoRa Mobility and Coverage Dataset (LoRaMC), a large dataset for LoRaWAN, which presents devices with mobility patterns in a city-scale and which aims at providing researchers with a means to develop new algorithms and assess their performance on common data, thus allowing for fair comparisons.

2 DATASET DESCRIPTION AND ORGANIZATION

LoRaMC comprises mobility traces of 3300 taxis integrating network coverage information related to 50 LoRaWAN GWs deployed throughout the operation in the metropolitan city of Rome. We start from the CRAWDAD Rome/taxi dataset [4] that includes location data of 320 taxis moving around the city and produce a new dataset with 3300 taxis that also include radio-level information, these were not present in CRAWDAD.

[1] https://www.iot-lab.info/

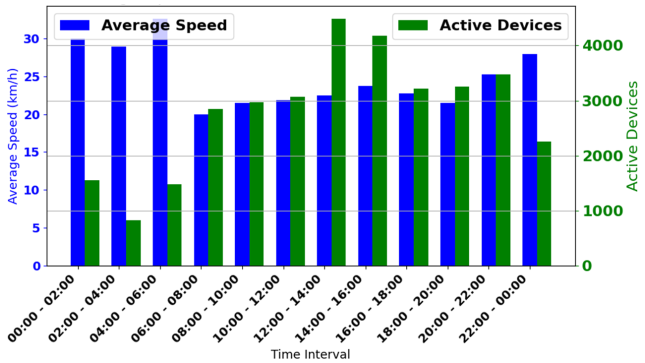

Figure 1: Trend of active devices and average speed per 2-hour time interval.

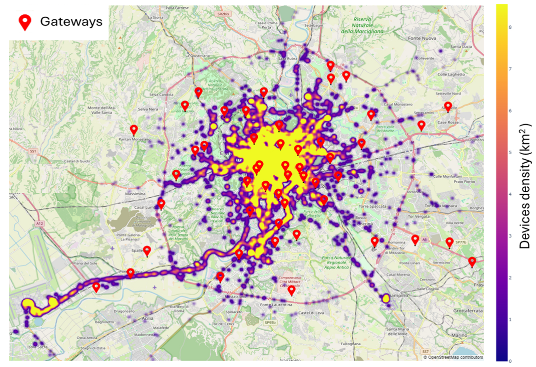

Figure 2: Dataset device density and coverage with 50 GWs over the entire period.

We first reduce the entire CRAWDAD Rome/taxi dataset into a single day by dividing the dataset into 30 different subsets of data, one for each day. This division is used to assign to each ID of the taxi also the day and now consider this new pair (ID, day) as the new identifier. Afterward, we projected all the tracks on only a single day, we divided each of these 3300 tracks into 288 "snapshots" of 5 minutes duration each. Snapshots represent the state of the taxi in that time window, for each taxi, if there is at least one position information in that specific time window of 5 minutes, we aggregate these locations in a single point representing the snapshot. Initial location resolution was 7 seconds, and complete location series are considered as part of the payload for each snapshot. If there is no data point regarding the position of a taxi in a snapshot interval, the cab is not included in the snapshot, as if it is inactive or the device is turned off, unable to send or receive any message.

Now there is the need to have another set of position data, containing information on a position related to the GWs, for this kind of information several options can be exploited: Decide arbitrarily the position of the GWs, set the GWs in a pre-defined schema as a grid, or decide the position of those based on the distribution of the cabs. Fig. 1 depicts the activity levels of taxis in terms of their quantity and speed. The figure is formatted as a histogram consisting of 12 bins, each representing a 2-hour interval. The height of each green bin indicates the number of active devices during that particular time interval, whereas the height of each blue bin represents the average speed of the active devices within that same time frame. Given that not all taxis are operational throughout the day, observe how the number of active devices fluctuates according to the time of day, with a corresponding increase during peak hours. In addition, the average speed of taxis is higher during the night, a time when traffic is typically reduced.

Given the location of the taxis during each snapshot, we used the ns-3 simulator [2] to simulate the radio coverage of a LoRaWAN deployment and include the resulting radio traffic information in the dataset. We positioned 50 GWs in a way such that they reflect a realistic deployment strategy as it would be designed by a LoRaWAN operator. Fig. 2 presents the density of the devices and the position of the GWs throughout the city area. Thus, for each snapshot, a separate simulation was carried out in which for each device present in that snapshot a total of 10 LoRaWAN frames were transmitted. For each frame, we monitor the potential reception, the number, and the positions of the GWs that have received that frame, extracting the power with which each GW receives the frames together with the data rate and the used channel/frequency for the frame transmission.

For each terminal, for each transmitted frame presented in the dataset, we include temperature, humidity, and pressure values as part of the payload. The final dataset output is structured in a single file, structured as a list of transmission events, one for each row, and for each row, we have a list of receptions, given the fact that a frame can be received from 1 or multime GWs or not received at all. More in detail, each row of the dataset has 8 (one is a list) fields that represent information about transmission, the position of the device, and the list that collects any receptions with their respective characteristics. The final dataset has the following structure: timestamp, lat, lon, high, deviceADDR, DataRate, FrameCounter, and ListOfGW. Elements of the ListOfGW list are: Timestamp, Sender deviceADDR, DataRate, FrameCounter, Frequency, RSSI (received signal strength indicator), GW MACaddress.

[2]https://www.nsnam.org/

ACKNOWLEDGMENT

This work was partially supported by the European Union - Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP E83C22004640001, partnership on "Telecommunications of the Future", PE00000001 - program "RESTART", and Investement 7 PE00000014 - CUP D33C22001300002, program SERICS.

REFERENCES

[1] Cisco's annual internet report highlights tool 2018-2023. In *CISCO*, 2018.
[2] LoRa Alliance. Lorawan 1.0.4 specification. https://lora-alliance.org/wp-content/uploads/2020/11/LoRaWAN-1.0.4-Specification-Package_0.zip, 2020.
[3] Cédric et al. Adjih. Fit iot-lab: A large scale open experimental iot testbed. In *IEEE World Forum on Internet of Things (IEEE WF-IoT)*, Milan, Italy, December 2015.
[4] Lorenzo Bracciale, Marco Bonola, Pierpaolo Loreti, Giuseppe Bianchi, Raul Amici, and Antonello Rabuffi. Crawdad roma/taxi, 2022.