Dataset: LoED: The LoRaWAN at the Edge Dataset

Laksh Bhatia
Imperial College London
London, UK
laksh.bhatia16@imperial.ac.uk

Michael Breza
Imperial College London
London, UK
michael.breza04@imperial.ac.uk

Ramona Marfievici
Digital Catapult
London, UK
ramona.marfievici@digicatapult.org.uk

Julie A. McCann
Imperial College London
London, UK
j.mccann@imperial.ac.uk

ABSTRACT

This paper presents the LoRaWAN at the Edge Dataset (LoED), an open LoRaWAN packet dataset collected at gateways. Real-world LoRaWAN datasets are important for repeatable sensor-network and communications research and evaluation as, if carefully collected, they provide realistic working assumptions. LoED data is collected from nine gateways over a four month period in a dense urban environment. The dataset contains packet header information and all physical layer properties reported by gateways such as the CRC, RSSI, SNR and spreading factor. Files are provided to analyse the data and get aggregated statistics. The dataset is available at: doi.org/10.5281/zenodo.4048321

CCS CONCEPTS

• **Networks** → **Link-layer protocols**; • **Computer systems organization** → Embedded and cyber-physical systems.

KEYWORDS

datasets, lorawan, gateways, urban environment

ACM Reference Format:
Laksh Bhatia, Michael Breza, Ramona Marfievici, and Julie A. McCann. 2020. Dataset: LoED: The LoRaWAN at the Edge Dataset. In *Proceedings of the Third Workshop on Data Acquisition To Analysis (DATA '20), November 16–19, 2020, Virtual Event, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3419016.3431491

1 INTRODUCTION

LoRaWAN is a wireless single-hop, long-range and Low-Power Wide Area Network (LPWAN). LoRaWAN has seen rapid adoption due to its communication coverage, ease of deployment, and simplified infrastructure management. Over 180 million LoRa-enabled devices[1] are already used for a variety of smart city and rural applications.

A significant body of academic work on LoRa and LoRaWAN has proposed improvements to data rate mechanism control [5] and solutions to decrease collisions and increase channel goodput [8], scheduling for reliable and efficient data collection [2], and scaling [7]. The validity and performance of the proposed improvements and solutions has been extensively tested in lab-like testbeds and simulations which have little in common with real-world environments and deployments. This is due to the absence of publicly available traces from the target environments. The IoT research community has recognised the relevance of quantitative evidence from real-world environments and deployments and has made available a few LoRaWAN datasets [1, 4, 6]. The fingerprinting dataset [1] contains over 120,000 traces collected over a three month period. Each trace stores the location of the device, the Received signal Strength Indicator sampled by the gateways upon receiving a packet, the spreading factor used, and the time of the received packet. The fingerprinting dataset provides no gateway capacity, load, or deployment environment information. The LoRa underground link dataset [6] is collected in an agricultural environment. It contains data from a low-density LoRaWAN deployment with only five transmitters and two receiver base stations.

In this paper we present LoED, the LoRaWAN at the Edge Dataset, which consists of traces gathered in central London in a mix of dense urban and park environments. Overall, 11,000,000 packets (referred to as packets in this paper) were collected at nine gateways during 2019 and 2020. We describe LoED in Section 2, and briefly discuss how the dataset can be exploited in Section 3.

2 DATASET

Setup. LoED was acquired from nine LoRaWAN gateways in central London. The gateway locations were representative of typical dense urban and park environments and cover different deployment conditions as shown in Table 1. Five outdoor gateways were deployed on the roof tops of large buildings, with a clear line-of-sight (LoS) to devices. Four indoor gateways were located near windows with limited LoS. One of the indoor gateway was placed on the ground floor of a college dormitory with no-LoS to any device. Each gateway forwarded received packets to a multiplexer which forwarded them to different Network Servers. Our server copied the packets and the gateway metadata to a time-series database. The gateway locations can be found in the dataset and table 1 contains information about the quantity of packets received at each gateway. We collected LoRaWAN packets over a 2-4 month period generated by smart city applications and research deployments. The gateways received from 5,000 up to 120,000 packets per day.

[1]https://www.semtech.com/lora

DATA '20, November 16–19, 2020, Virtual Event, Japan

Gateway ID	Description of location	Number of days	Total packets	max packets in a day	avg packets per day
00000f0c210281c4	Dense outdoor area, on top of a building	19	1326687	82781	69826
00000f0c22433141	Roof of a low building in a non-dense area	36	144777	6579	4022
00000f0c210721f2	Top of a building in a very dense area and large open spaces	56	5757575	121368	102814
00000f0c224331c4	Indoor in the ground floor a building, surrounded by buildings	15	17029	1625	1135
00800000a0001914	Deployed inside a university building	573	76706	2366	134
00800000a0001793	Deployed inside a university building	552	186592	9596	338
00800000a0001794	Deployed inside a university building	17	61080	4810	3593
7276ff002e062804	Deployed on top of a tall university building, with large open spaces	131	1201916	15254	9175
0000024b0b031c97	Urban area, top of building, dense deployment	131	2490639	25708	19013

Table 1: Gateway data and location overview.

Overall, 11, 263, 001 packets from 8, 503 unique device addresses with valid Cyclic Redundancy Check (CRC) and an uplink packet type were collected including packets with failed CRC.

The Data. We assumed that all collected packets were using the explicit header (EH) mode which enabled metadata extraction of: the device address, packet type, counter value, port number and other information like ADR. Upon packet reception at the gateway the timestamp, physical layer CRC status, frequency, spreading factor, bandwidth, coding rate, sampled RSSI and Signal to Noise (SNR) values are added to the packet information. Each trace consists of the following fields:

time	Time at which the packet was received
physical_payload	Raw payload received
gateway	Gateway where the packet was received
crc_status	Physical layer CRC
frequency	Radio frequency of the packet
spreading_factor	LoRa spreading factor of the packet
bandwidth	Bandwidth of the received packet
code_rate	LoRa coding rate of the packet
rssi	Sampled RSSI value at packet reception
snr	Sampled SNR value at packet reception
device_address	Device address for the packet
mtype	Packet type
fcnt	Counter value of the packet
fport	Port of the packet

The LoED dataset is publicly available at [3] and includes: *i*) all packets received at the nine gateways, *ii*) one pre-processed CSV data file for every day of the collection campaign. The files are saved in **dd_mm_yyyy.csv** format, *iii*) a set of scripts for processing and plotting the data and, *iv*) a preliminary analysis of the data.

Preliminary insights. LoED exposes insights into how LoRaWAN operates in real-world urban deployments and provides data such as: *a*) number of packets per day at a gateway, *b*) total number of packets per node, *c*) distribution of different packets types at gateways, *d*) distribution of frequencies used at gateways, *e*) distribution of spreading factors used at gateways, *f*) distribution of RSSI values at a gateway, *g*) distribution of SNR values at a gateway. In Figure 1 we see the LoRa spreading factor usage at each gateway. Of note are the dominate use of spreading factors: 7, 8 (at two of the gateways) and 12. This may increase the probability of collisions as the number of devices using the same spreading factor rises.

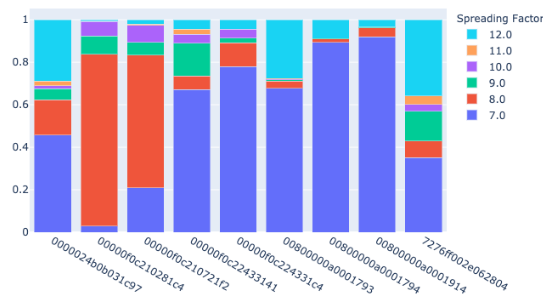

Figure 1: Distribution of spreading factor at each gateway.

As a consequence, the performance of the applications running on the devices in the network may decrease and require further, lower-level, investigation.

3 DISCUSSION

The LoED dataset can be used by the LoRaWAN community for many purposes. It can be used to characterise parameter usage and highlight long-term trends for LoRaWAN applications and devices in an urban environment. LoED can provide test data to inform the design of scheduling algorithms and protocols, to ensure they are well-suited to the target applications and environments. Further, LoED can be plugged into capacity planning systems, which use different statistical and machine learning, to derive optimal parameters to improve network throughput, reduce interference or determine locations for new gateways to improve coverage.

REFERENCES

[1] Michiel Aernouts et al. 2019. Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas. https://doi.org/10.5281/zenodo.3342253
[2] Laksh Bhatia et al. 2020. Control Communication Co-Design for Wide Area Cyber-Physical Systems. arXiv:2008.07492 [eess.SY]
[3] Laksh Bhatia, Michael Breza, Ramona Marfievici, and Julie A. McCann. 2020. LoED: The LoRaWAN at the Edge Dataset. https://doi.org/10.5281/zenodo.4048321
[4] Norbert Blenn and Fernando Kuipers. 2017. LoRaWAN in the Wild: Measurements from The Things Network. arXiv:1706.03086 [cs.NI]
[5] Martin Bor and Utz Roedig. 2017. LoRa Transmission Parameter Selection. In *Proc. of DCOSS*.
[6] Rachel Cardell-Oliver et al. 2019. Dataset: LoRa Underground Farm Sensor Network. In *Proc. of DATA*.
[7] J. Haxhibeqiri et al. 2019. Low Overhead Scheduling of LoRa Transmissions for Improved Scalability. *IEEE Internet of Things Journal* 6, 2 (2019), 3097–3109.
[8] N. Kouvelas et al. 2020. p-CARMA: Politely Scaling LoRaWAN. In *Proc. of EWSN*.