2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) | 978-1-6654-0424-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/PERCOMWORKSHOPS51409.2021.9431070

Artifact: The Feasibility of Dense Indoor LoRaWAN Towards Passively Sensing Human Presence

Jascha Grübel*, Tyler Thrash†, Didier Hélal‡, Robert W. Sumner*, Christoph Hölscher*, Victor R. Schinazi§

* ETH Zürich, Zürich, Switzerland
Email: jascha.gruebel@inf.ethz.ch, robert.sumner@inf.ethz.ch, christoph.hoelscher@gess.ethz.ch
† Miami University, Oxford, OH, USA
Email: thrashst@miamioh.edu
‡Orbiwise, Geneva, Switzerland
Email: didier.helal@orbiwise.com
§ Bond University, Robina, Australia
Email: victor_schinazi@bond.edu.au

I. Introduction

We present a large data set of 14 million entries of indoor Long Range Wide Area Network (LoRaWAN) transmission metadata to study Dense Indoor Sensor Networks (DISN) [1]. We also provide a mixed model implementation in an R-Markdown file for analyzing the transmissions based on the metadata to obtain all transmission-related results of the main paper. The data set will regularly expand as the system continues to collect data until at least 2022.

Understanding Long Range Wide Area Network (LoRaWAN) transmissions indoors is critical to test the feasibility of Dense Indoor Sensor Networks (DISN). While LoRaWAN has been widely studied outdoors, we believe that indoor setups require a different conceptualization to account for signal shadowing and multi-path fading. The data set has been used in the accompanying paper to compute an indoor model of LoRaWAN transmissions. We provide the transmission data to enable other researchers to develop alternative indoor models for comparisons. The data set consists of all the LoRaWAN transmissions from the first six months of the prototype of the digital twin described in the paper. The data has been exported from the PostgreSQL data base on the EVEREST server described in Section III-B of the manuscript. The annotated transmission data was supplied by the LoRaWAN network server via webhook and stored on our server. The data has been stripped of identifying information with respect to the building because of privacy concerns.

II. Data Description

The public data set consists of a single table containing 21 characteristics of 14 million received transmissions (see Table I). Not included in the public data set are geometric relations between sensors and gateways because of privacy concerns.

In the data set, each row represents the reception of transmissions on the network server and is uniquely identified by the combination of the identifying variables `transmission`, `device`, and `gateway`. Under LoRaWAN, a transmission of a device can be received by any gateway, and therefore, a single transmission may appear up to 3 times, once per gateway. In some scenarios, there may be less transmission due to environmental (e.g., wall material) or network factors (e.g., gateway up-time).

The devices remained in the same location during the 6 months of data collection and have several immutable characteristics. The `device_type` corresponds to the 5 sensor types presented in Table II of the section III-A of the paper. The `distance` between gateway and device was computed with the PostGIS extension. The device locations were predetermined during installation along with a building information model (BIM). Similarly, the gateway floor `gw_floor` and device floor `dev_floor` were annotated during installation, and the floor distance `floor_dist` was directly computed from the difference between the two. We only used `floor_dist` for modeling, but the other two variables can be used for visualization. The BIM indicates that a floor has a height of 3.36 m which was used to calculate maximal transmission distances.

Transmission characteristics vary with each transmission and describe the LoRa configuration, network meta data on the signal strength, and varying device properties. The LoRa configuration is described in the Data Rate (DR) `dr`, the Spreading Factor (SF) `sf`, the frequency `freq`, and the Time-on-Air (ToA) `toa`. The network metadata consists of the measured `snr` and `rssi`, the arrival time `time`, and a Boolean flag `main_gw` that indicates which was the main recipient gateway with the strongest received signal. The device property that is transmission-specific is the `frame_count`, indicating the internal ID of the transmission within the device. The frame count allows us to determine the Frame Error Rate (FER). A gap in the frame count indicates that the device has sent a transmission that was not received by the network.

The derived measures determine two types of concurrent transmissions. First, multi-channel concurrent transmissions `concurrent` occur at the same time but on a different channel and SF and have no impact on the transmission quality. Second, destructive concurrent transmissions

TABLE I
COLUMN DESCRIPTION FOR THE DATA FRAME OF LORAWAN TRANSMISSIONS.

Variable	Description	Data Type	Unit	Range
Identifying variables				
transmission	ID of transmission	Integer	ID	2–6645488
device	ID of device	Integer	ID	1–136 & 140–393
gateway	ID of gateway	Integer	ID	137–139
Device characteristics				
device_type	main sensor type	Integer	ID	1=Orbiwise noise; 2=ERS PIR; 3=ERS CO2; 4=Browan PIR; 5 = Browan VOC
distance	between gateway and device	Float	m	0.8062–64.4224
dev_floor	Floor of device	Integer	Descriptive	0–7
gw_floor	Floor of gateway	Integer	Descriptive	0–7
floor_dist	Number of floors between gateway and device	Integer	Count	0–7
Transmission characteristics				
dr	Data rate used	String	Descriptive	"SF[7–12]BW125"
sf	Spreading factor used	Integer	Descriptive	7–12
freq	Frequency used	Integer	Hz	867100000–868500000
toa	Time-on-air	Integer	ms	57–2138
snr	gateway measure of SNR	Float	db	-21.0–14.0
rssi	gateway measure of RSSI	Float	dbm	-131.0–123.0
time	arrival time at gateway	String	ISO 8601 Date & Time	2020-02-25 14:39:00–2020-09-10 12:33:14
main_gw	main gateway that received the data	Boolean	Logical	True–False
frame_count	Sensor-based frame counter	Integer	Count	1–56739
Derived measures				
interval_start	Equal to time minus toa	String	ISO 8601 Date & Time	2020-02-25 14:39:00–2020-09-10 12:33:14
interval_end	Equal to time	String	ISO 8601 Date & Time	2020-02-25 14:39:00–2020-09-10 12:33:14
interval_start_buffer	Start interval with 25% buffer	String	ISO 8601 Date & Time	2020-02-25 14:39:00–2020-09-10 12:33:14
interval_end_buffer	End interval with 25% buffer	String	ISO 8601 Date & Time	2020-02-25 14:39:00–2020-09-10 12:33:14
concurrent	Other transmissions that do not impact this transmission	Integer	Count	0–29
concurrent_destructive	Other transmissions that could cancel this transmissions	Integer	Count	0–5

concurrent_destructive on the same frequency and SF make the reception of all transmissions at a single gateway impossible. However, the physical distance combined with short ToA make it possible that a destructive concurrent transmission at one gateway is non-concurrent at another. We employ multiple gateways to detect destructive concurrent transmissions. First, the transmission interval is based on the time and toa described with the variables interval_start and interval_end. Since our gateways are network synchronized, we must allow for a delay (up to 20 ms). We use a conservative estimate of 25% of toa to buffer the interval on both ends described in interval_start_buffer and interval_end_buffer. The buffered intervals are used to compute concurrent and concurrent_destructive.

III. DATA AVAILABILITY

The data set is accessible on Zenodo as a CSV file: https://doi.org/10.5281/zenodo.4476317. Zenodo is a service of CERN that allows free upload of scientific data. A single submission (up to 50GB) is never deleted as long as CERN is operating. Currently, this promises data availability at least for the next 20 years with a high probability that the funding agencies behind CERN will extend the operation indefinitely.

We withhold some information because of privacy concerns. For more information, please contact the first author.

IV. RESULT VALIDATION

In addition, the Zenodo repository contains an R Markdown file to reproduce the model and model predictions in the main paper. At least 16GB RAM are required to run the models.

1) Install and open RStudio (https://rstudio.com/products/rstudio/download/).
2) Open 2021_PerCom_mixed_model.Rmd.
3) Ensure that LoRaWAN_DISN_transmission_meta_data.csv is in the same folder.
4) Press Ctrl+Alt+R to run the code.

All models and graphs from the main paper are reproduced as shown in the knitr output files of the above steps: 2021_PerCom_mixed_model.[pdf/html].

ACKNOWLEDGMENT

This research is funded by grant ETH-15 16-2 and supported by OrbiWise with hard- and software. We thank Michal Gath-Morad for the BIM.

REFERENCES

[1] J. Grübel, T. Thrash, D. Hélal, R. W. Sumner, C. Hölscher, and V. R. Schinazi, "The feasibility of dense indoor lorawan towards passively sensing human presence," in *2021 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2021, pp. 1–2.