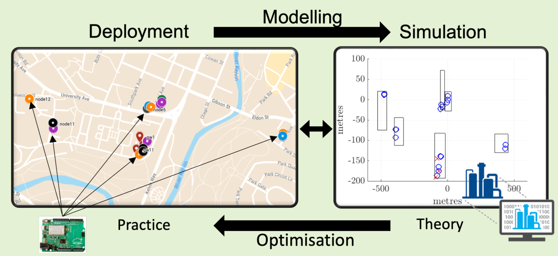
Comparative Analysis of an Urban LoRaWAN Deployment: Real World vs. Simulation

Bruno Citoni, Shuja Ansari, Qammer H. Abbasi, Muhammad A. Imran, Sajjad Hussain

***Abstract*—LoRaWAN simulations are a flexible way to analyse the behaviour of this LPWAN technology in scenarios that are unfeasible to deploy due to their scale and the number of devices required. Parallel to this, there is also a continued lack of larger-scale LoRaWAN deployments in current literature. Crucially, none of these studies involves comparison with any theoretical model, such as discrete-time simulation or mathematical analysis, for validation. In this paper we deploy a 20 nodes LoRaWAN network around the University of Glasgow's campus, analyse the results and then proceed to develop an NS-3 simulation to recreate and match as faithfully as possible the behaviour and topology of the physical deployment. The performance of both the deployment and the simulation is then compared, and the results show that while the complexity of the simulation is kept relatively low, it is possible to get simulation results within about 20% of the deployment results.**



***Index Terms*—Deployment, IoT, LoRaWAN, NS3, PDR, Real-life, Simulation.**

I. Introduction

THE term "Internet of Things" (IoT) refers to distributed measurement systems made up of wirelessly connected end devices, transmitting sensing data and receiving feedback from a central processing unit. Over the past decade, with the increased number of connected devices and sensors being developed, we have witnessed the rise of IoT technology in both consumer applications like smart-health and smart home as well as over multiple industries, such as agriculture, automotive, oil and gas. Such networks are based on a communication technology that is low in power consumption, has a high maximum range, and is highly scalable, capable of dealing with many wirelessly interconnected devices at the same time. These requirements are not fully met by any of the traditional communication protocols, be they short-range, like Bluetooth or Wi-Fi, or long-range, such as cellular and satellite. LPWANs (Low Power Wide Area Networks) are protocols designed to complement the existing ones and to better fit the IoT ecosystem specifically. LoRa (from Long Range), NB-IoT, Sigfox, and Weightless are all examples of LPWANs used in IoT and IIoT (Industrial Internet of Things) applications.

In particular, LoRaWAN, a MAC layer protocol built on top of the LoRa proprietary modulation technique, has received much attention in the past few years. This is partly due to its open-source nature, the promise of low capital and maintenance costs, and potential long battery life with very high communication range. Moreover, LoRaWAN is very flexible, allowing for both private and public networks, and affording the designer control over its performance via multiple trade-offs, such as increasing the communication range at the expense of battery life. This makes it suitable for a variety of specific use cases, since multiple parameters can be fine-tuned. The initial release of LoRaWAN dates back to 2015, and yet there is still a lack of performance analysis studies based on real-life deployments or scalable testbeds, particularly large-scale ones, with a few notable exceptions [1], [2]. For convenience, such networks are mainly analysed through discrete-time simulations [3] or equivalent mathematical models [4], [5]. More so, to the best of our knowledge, no previous work tries to recreate in simulation, as faithfully as possible, the same real-life deployment carried out, either working entirely empirically or theoretically. While real, large-scale networks will always be hard to deploy and analyse, the following work aims to provide a metric of just how effective simulations can be so that they can be more confidently used instead, for LoRa as well as other technologies.

In line with the gaps identified in the related work (Section III), the key contributions of this work are:

- Deployment of a 20-node network on the campus of the University of Glasgow. These nodes operate with one of five different transmission behaviours to make the deployment closer to what a real-life application would be, with different sensors having different uplink schedules.
- Porting of the entire physical deployment scenario as close to reality as possible to NS-3, a discrete-time simulator

This work was supported by the main author's UK EPSRC PhD studentship through grant EP/R512266/1

B. Citoni, S. Ansari, Q.H. Abbasi, M.A. Imran and S. Hussain are with the James Watt School of Engineering, University of Glasgow, Glasgow, UK;email: b.citoni.1@research.gla.ac.uk.

- Analysis of the results of both methods while identifying points of conflict between the two, highlighting further improvement needed to the discrete-time simulation routines and lessons learnt.

The rest of this paper is structured as follows. First, we briefly discuss the characteristics of LoRaWAN that will be relevant during the rest of the paper in Section II. We then look at the relevant literature in Section III before describing the setup and results of the physical deployment in Section IV. Then we describe the setup and results of the simulation, as well as checking if and how the results from both methods are comparable in Section V. Conclusions and ideas for further work are given in Section VI.

II. LoRa and LoRaWAN

A. LoRa physical layer

LoRa is a proprietary chirp spread spectrum (CSS) modulation technique patented by Semtech. It is designed to enable communication in the range of kilometres. To achieve this, a number of configurable radio parameters that control the maximum achievable communication range, power consumption, and data rate can be tuned.

These parameters are the bandwidth (BW), which is the spectrum over which the modulated signal is spread, the coding rate (CR), which is used to perform forward error correction techniques, and the spreading factor (SF) of the transmission. An example of the control one can have over the network performance is how the sensitivity of the receiver can be improved by changing the SF of a transmission, which can range from 7 to 12. Increasing SF will boost coverage while reducing the bitrate and increasing the packet's Time-on-Air (ToA), and vice versa.

The chance of a packet being received and decoded accurately is strongly correlated with its ToA. While raising SF and therefore the ToA makes it easier for the packet to be received by a gateway by increasing its range, it also increases the likelihood of collisions with other packets, saturating a receiver's demodulation channels, and exceeding the permissible duty cycle limitations.

LoRa usually operates in the license-free, region-dependent ISM frequency bands, 863-870 MHz for Europe and 902-928 MHz for the United States, although it can also function in the lower ISM bands 433 MHz and 169 MHz.

Using the unlicensed ISM frequency spectrum reduces the cost of deployment, while also limiting the highest attainable data rate due to constraints on available air-time per device on the same frequency. For the regularly utilised frequency sub-bands of 863.00-868.00 MHz and 868.00-868.60 MHz, the duty cycle is imposed at 1%. This implies that a device can only send data on each of the two sub-bands 1% of the time [6].

B. LoRaWAN MAC layer

LoRaWAN is the open-source MAC protocol built on top of the LoRa modulation and managed by the LoRa Alliance, an open, non-profit association aiming to promote the adoption of the LoRaWAN standard.

LoRaWAN networks contain three main elements:

- Nodes or End Devices (ED) are sensor boards responsible for collecting data or implementing instructions via actuators through LoRa-based communication.
- Gateways (GW) are devices that forward packets coming from nodes to a network server and vice versa, acting as a logically invisible bridge between nodes and the network.
- The Network Server (NS), which handles the deduplication of received packets, rejection of corrupted/unwanted packets, and scheduling messages to be sent to specific nodes through gateways in range.

A typical LoRaWAN network is organised in a star-of-stars topology, with nodes transmitting to all gateways in range rather than having a direct link to any single gateway. Uplink allows devices to communicate with the network, while downlink allows them to receive data. Downlink data to a node must always be routed through a gateway in range because standard LoRaWAN does not provide native node-to-node communication, though this can be implemented via different MAC layers built on LoRa modulation. Downlink communications must be sparse since gateways are subject to the same duty cycle limits as nodes. Downlink also affects the overall Quality-of-Service (QoS) of a network by adding additional interference and preventing gateways from receiving uplink messages while sending in downlink.

LoRaWAN networks define three different classes for its EDs, Class A, Class B and Class C. Class A devices provide bi-directional communication, but downlink can only be received over two brief time windows being opened after an uplink is sent. Class B provides extra slots for downlink communication at periodic times following scheduling via beacons. Class C devices continuously listen for downlink messages, except when transmitting [7]. While research has been extended to other classes, Class A remains the most popular and used of the three.

III. Related works

Modelling existing network deployments is an established practise in literature for a number of different communication technologies and via different simulators [8], [9]. However, the practise has not been widely extended to LoRa and LoRaWAN yet, with existing literature either focusing on simulation or empirical deployments. The work in [10] reports the RSSI (Received Signal Strength Indicator) values obtained in a three-storey building in Hyderabad, India. A single node is placed in different sections of the building and information is gathered by a single stationary gateway. This investigation is meant to provide insight on how the path loss for LoRaWAN is affected by indoor obstacles like walls and doors. Similar work is presented in [11] where the authors report the measured results of a field experiment across multiple locations in Japan. Once again the packets are only used to validate the empirically obtained path loss compared to the Okumura-Hata model. The authors in [12] perform an indoor and outdoor coverage analysis in a Smart Farm scenario. Together with information on RSSI and SNR (Signal-to-Noise Ratio) of the received packets, they also report the PDR (Packet Delivery

Ratio) achieved by a node in different locations against the distance from a common gateway.

In [13], the authors present an analysis of a testbed deployed in the University of Calgary in Canada. Two custom gateways are placed indoor on the 7th floor of a building, with 4 nodes operating on different frequencies in the 915 MHz band. PDR for each floor of the building as well as for larger outdoor distances are recorded. Additional experiments were carried out to characterise how different parameters such as different spreading factors and different packet size can affect the PDR.

In their study, Yasmin et al. [2] deployed 331 nodes, each with 5 different sensors at the University of Oulu. They are all set to have a fixed transmission interval of 900 seconds, packets with constant length of 24 bytes, operating with a spreading factor of 7. The Adaptive Data Rate (ADR) algorithm is off and, since the devices will not require confirmation of packet receipt, there is theoretically no downlink. Data received at a single gateway are collected and analysed in terms of RSSI, SNR, and PDR and show that no device was below 25% PDR and some had PDR $>$99.5%.

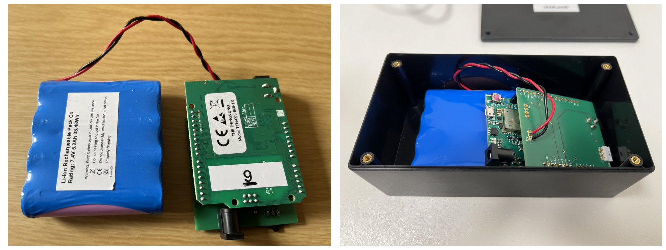
The work in [1] studies the results obtained from a large deployment in Shangai, with data collected over 8 months. They investigated the packet loss rate (defined as the opposite of PDR) and tried to find some correlation between various parameters such as the distance between transmitter and receiver and that of the RSSI and the SNR with the PLR (packet loss rate). Two different types of devices with different transmitting behaviours are included in the study, and final considerations on the results are brought forward in terms of the quantity of nodes in an area, the gateway deployment strategy, the different transmitting schedules of the two types of devices, and the impact of packet collisions in the network.

Finally, in [14], the authors carry out a similar analysis in Brno, Czech Republic, over an area of 288 km^2. More than 20 gateways are present throughout the city, and a single device is cycled through 231 test locations, while operating at spreading factor 12 and on the mandatory EU frequency bands. Timestamps RSSI, SNR, and reached GW locations were recovered from the network server and analysed in MATLAB. The overall PDR throughout the experiment is 83%, and some very interesting results are also recorded in terms of RSSI, PDR, ADR and gateway positions.

IV. Physical Deployment

A. System Model

A total of 20 end devices were assembled for this study. Each device is made up of a The Things UNO [15] and a custom PCB shield used to connect a 7.4V, 5.2Ah battery pack while providing a facile way to recharge it, shown in Fig. 1. The devices were placed in plastic enclosures and scattered around the University of Glasgow's 70-acre wide Gilmorehill campus in several indoor locations. These were varied in altitudes and room usage, but can be overall regarded as office spaces. None of the nodes were moved throughout the experiment, although external changes in the rooms and buildings were impossible to track. According to previous testing and calculations, each device was supposed to last on a single battery charge for approximately 3 weeks.



(a) (b)

Fig. 1: One of the deployed nodes and its plastic enclosure.

Each node is assigned an ID from 1 to 20 and one of five possible transmission routines, hereby referred as Tx Groups. These are meant to diversify the deployment and mimic how in a real-life network there might be different update needs for different data features. For instance, a node tasked with recording temperature in a warehouse may need to be heard from once every 10 minutes, while a light intensity sensor only needs to transmit data a couple of times during the whole night. As far as we are aware, the idea of diversifying the operation of nodes not only based on their data rate, but on their operation and application, has not been considered before. In related works, nodes are instead generally programmed or set in simulation to have fixed, periodic schedules across the entire network.

Devices in Tx Group 1 (ID 1,6,11,16) are programmed to transmit every 300 seconds, those in Tx Group 2 (ID 2,12,17) transmit as fast as the currently used SF will allow. For devices in Tx Group 3 (ID 3,8,13,18), a random interval between 300 and 900 seconds is chosen for each transmission, and for Tx Group 4 (ID 4,7,9,14,19) a random interval between the fastest the current SF will allow and three times that same value. Finally, devices in Tx Group 5 (ID 5,10,15,20) will transmit every 600 seconds for approximately 12 hours (representing operation during daytime), and then transmit roughly every 7200 seconds for 12 hours (representing operation during nighttime) before going back to daytime operation.

In addition to the three gateways (ID 1,2,11) we deployed in the James Watt South building, we were expecting to reach a number of third-party gateways, as the network used in our work is managed by IoT Scotland and widely used throughout the country. Gateways 1 and 2 were placed indoor, while Gateway 11 on the roof of the same building. Analysis of the results confirms that a total of 15 different gateways received at least one packet from any of our devices. We depend on the accuracy of the location provided by the unknown third-parties operating these gateways, as we will use the GPS metadata recovered from our packets to place the gateways in space when recreating the network in simulation. The geographical extent of the network is visible in Fig. 2. For a more detailed view, a custom, interactable map is available[1].

All nodes start transmitting using SF12, packets with 10 bytes of payload. They are set to hop randomly between 8 possible frequency channels: the mandatory channels at 868.1, 868.3, 868.5 MHz and five additional ones at 867.1, 867.3,

[1]https://tinyurl.com/2p9npexz

867.5, 867.7, 867.9 MHz. The ADR algorithm is enabled both on the nodes and on the network server, which means that the SF and the transmitting power of each node can be increased or decreased throughout the deployment life cycle. While the on-device side is formally standardised, the network-side algorithm is developed by each network operator based only on recommendations [16]. The algorithm employed in both the chosen network and the simulation for this study are supposedly the same, but as this cannot be confirmed for certain, it could lead to some discrepancy in results.

The experiment lasted for a month, with nodes placed between the 6th and the 8th of December 2021 and picked up in January 2022, when it was noticed that no more data was being collected, as all nodes ran out of battery within approximately 3.5 weeks.

B. Results

Key metric to quantify the QoS of a LoRaWAN network found in literature is its PDR or its inverse, Packet Loss Rate (PLR). RSSI, SNR and overall throughput have also been used [1], [2], [6], [17]. When gathering and analysing results from a real LoRaWAN deployment, the only information that can be studied is the one coming from the packets that have been correctly received. While one can extrapolate the behaviour of a device based on information such as the frame-counter of its last received transmission, there is no guarantee that the packet was the last transmitted by the node, but only that it was the last one that was received. The consequence of this, among other things, is uncertainty regarding key factors such as a the PDR of a network. This is in contrast with simulations, where all packet related events can be saved and compared to what it is actually received at the network server.

Fig. 2: Geographical scale of the deployed network

TABLE I: Nodes ID, groups, operating time, packets sent, delivered and PDR.

Node ID	Tx Group	Operating hours	Sent	Delivered	PDR (%)
1	1	597	7055	5965	84.55
2	2	42	11997	10313	85.96
3	3	557	3308	2961	89.51
4	4	262	5584	4139	74.12
5	5	511	1573	1227	78
6	1	98	1122	876	78.07
7	4	231	56146	55911	99.58
8	3	533	3178	3150	99.12
9	4	13	2684	2470	92.03
10	5	497	1544	1539	99.68
11	1	589	6958	4259	61.21
12	2	512	20837	6954	33.37
13	3	514	2913	1828	62.75
14	4	7	158	46	29.11
15	5	496	1532	590	38.51
16	1	23	280	266	95
17	2	355	97754	97584	99.83
18	3	529	3191	627	19.65
19	4	321	6827	4749	69.56
20	5	79	248	248	100
TOTAL			234889	205702	87.57

As it is shown in Table I, the devices we deployed in the network experienced very different operating times, even when powered with the same, fully charged batteries. These operating times are based on the timestamp of the first and last packet received by each node and can be assumed to vary for a number of reasons.

Node 14, during the 7 hours was certainly alive only delivered 46 packets out of the 158 that its frame counter suggests were attempted. It probably failed to maintain the connection alive due to interference, or a miscalculated change in data rate from the ADR that was not subsequently fixed. The outage of other nodes like 20 and 9, both of which were performing very well during their time alive, could be explained by changes in the building, or even room condition.

The initial effort was to mitigate the impact of these unexpected discrepancies on the accuracy of the PDR of the nodes and determine how long each device was actually supposed to last. The longest time alive for nodes within each Tx Group was taken as the expected for the whole group, assuming the same operating time due to the same transmitting schedule.

However, as mentioned, Tx Groups 2 and 4 have the ability to set their transmission interval dynamically, based on their SF. This ability, coupled with the fact that the SF of each device can be changed because of the ADR algorithm, gives rise to devices operating in the same group having greatly different life expectancies, such as node 12 and node 17. This effect is not as prominent for Groups 1, 3 and 5, which have a static, SF independent Tx Interval, and where the difference in operating time is probably due to the reasons mentioned above.

Because of this, it was decided to work with the best-case scenario instead. Devices are considered to be operating only during the time they were active for certain, and the simulation is adapted to have each device run just as long. The resulting PDR will likely not be as accurate a reflection of the Quality-of-Service and throughput of the deployed network.

TABLE II: Total number of packets successfully received for each gateway-node pair, including duplicates.

	GW 1	GW 2	GW 3	GW 4	GW 5	GW 6	GW 7	GW 8	GW 9	GW 10	GW 11	GW 12	GW 13	GW 14	GW 15
Node 1	341	0	0	0	0	0	0	21	0	0	5920	0	0	0	0
Node 2	475	0	0	0	0	0	0	23	0	0	10271	0	0	13	0
Node 3	831	0	0	0	0	0	3	34	0	0	2848	0	0	62	0
Node 4	0	0	0	0	0	0	0	1789	0	0	3533	0	0	2716	0
Node 5	6	0	0	0	0	0	0	18	0	0	1224	0	0	0	0
Node 6	28	0	0	0	0	0	0	31	0	0	872	0	0	3	0
Node 7	1	54459	0	0	0	10	10	753	0	0	54358	0	0	3	0
Node 8	2457	269	0	0	0	0	0	1869	0	0	2956	0	0	402	0
Node 9	58	1	0	0	0	0	0	19	0	0	2468	0	0	15	0
Node 10	1499	1401	0	0	0	0	0	271	0	0	1472	0	0	93	0
Node 11	0	0	0	0	0	0	0	0	0	0	4259	0	0	11	0
Node 12	0	0	0	0	0	0	0	0	0	0	6785	0	0	1278	0
Node 13	0	0	0	0	0	0	6	0	0	0	1815	0	0	228	0
Node 14	0	0	0	0	0	0	0	41	0	0	1	0	0	13	0
Node 15	0	0	0	0	0	0	0	0	0	0	590	0	0	0	0
Node 16	0	0	0	0	0	0	0	114	0	0	69	0	0	256	0
Node 17	28172	95931	1409	6	206	202	1179	8853	174	1072	95111	3613	20	680	9
Node 18	0	0	0	0	0	0	0	0	0	0	627	0	0	0	0
Node 19	0	0	0	0	0	0	1034	0	0	0	4498	0	3	284	0
Node 20	161	244	172	1	71	9	182	131	58	125	239	216	57	99	0

One should account for the fact that a device should have lasted longer and taking all packets that should have been sent during this period as lost. However, restricting the simulation to the best-case scenario ensures the most comparable results.

Together with the standard LoRaWAN frame counter, increasing with every transmission and included in the mandatory packet header, we also added a firmware counter in the payload, called *MyCounter*, increasing on the device each time a packet was sent to the send routine on-board. Discrepancies between *MyCounter* and the regular frame-counter can then be used to check how many packets were stopped on the device because of duty cycle restrictions or other reasons.

Taking the *MyCounter* value of the last packet received by a node as the number of packets sent by that node and taking the recorded instances of data from that device arriving at the network server as the number of packets received will yield a PDR for each node based on:

$$PDR = \frac{Packets\ received}{Packets\ sent} \times 100 \tag{1}$$

The results are shown in Table I.

On average, the groups with the highest PDR are those with the slowest Tx Interval. This is not as noticeable here given the fact that only 20 devices are being operated concurrently (plus any other outside our test network that we have no knowledge of). The more frequently a node sends, the more likely it is that its packets will collide in the air with others, but as this is an effect that scales with the size of the network is not as prominent here. However, we can see that Tx Groups 1 and 5 have 7% higher PDR than Tx Groups 2 and 4, from Fig. 3.

Table II shows the amount of packets received by each gateway from each of the nodes in the deployment. This includes packets that have not yet been deduplicated by the network server, and as such the total received from each node is expected to be higher than the number of packets received, as most will likely be received by more than a single gateway. The network as a whole is heavily influenced by mainly 2 devices: gateway 11, receiving the most messages, and the only one receiving at least once from every node, and node 17, which similarly, delivers the most messages and is the only node which delivers at least once to every gateway. Most importantly, this table highlights the impact of gateway and node positioning. Node 17 and gateway 2 are the pair with the most packets exchanged between them in the uplink direction. They are also the closest, being only 1 metre apart. Most of the third-party gateways to which our nodes made connections are in the south of the University Campus, and Node 17 is only the second southernmost device. Node 7 is technically the southernmost deployed node, yet it does not enjoy the same high connectivity that devices 17 and 20 did. This is likely because, despite being placed on the same floor and in the same building, node 17 and 20 were positioned in a larger, emptier room with big windows, while node 7 is on a desk in a much smaller, busier and less ventilated room.

In a similar vein, node 14, the one that lasted the least, with only 7 active hours recorded, preferred connection to gateways 8 and 14 rather than gateway 11. This is particularly interesting as between node 14 and gateway 11 there is Kelvingrove Park, an open space with no buildings, while between node 14 and gateways 8 and 14 there is sandstone tenements and the expected urban sprawl. This possibly suggests that the conditions of the room and adjacent areas in which device 14 was placed prevented connection to gateway 11. This is also confirmed by the fact that nodes 4 and 16, placed in the same building as node 14, had a similar experience, though crucially they had a better overall connection with gateway 11.

V. Network Simulation and Comparison

A. System Model

The simulation is modelled and performed using the discrete-time network simulator NS-3, running a modified version of the modules first developed by Magrin et al. [3] for LoRaWAN analysis. These modules, along with NS-3, are widely used in literature and currently among the best and most complete simulating tool for LoRa and LoRaWAN behaviour [18].

Nodes and gateways are placed in the 3D space based on the coordinates gathered from the deployment as well as from the network server. Approximations of buildings containing nodes and gateways are also placed in the simulated space. The resulting network topology is comparable to the actual deployed network as described by the map in Fig. 2. ADR is present in its standard form, as described in the deployment setup.

The core of the simulation is represented by the channel model. For this study, choices fell onto the Log-Distance, Okumura-Hata and Friis models. All of these have previously been used to describe channel loss in LoRaWAN simulations, mathematical models, and analysis based on empirical data [3], [19]–[22]. For the Log-Distance case, we used different coefficients for γ and PL_0, based on the same formula for urban environments used in [3]. Similarly to [23] and [24] we also performed linear regression of the obtained RSSI values from the deployment and obtained γ and PL_0 coefficients by using the following relation:

$$RSSI = P_{tx} - PL + G_{tx} + G_{rx} \tag{2}$$

with

$$PL = PL_0 + 10\gamma log_{10}(d/d_0) + X_g \tag{3}$$

Neglecting the antenna gains G_{tx} and G_{rx} and the fading factor X_g (included by other means in our channel model), and with the standard transmission power of 14 dBm and d_0 of 1 metre we get:

$$RSSI = 14 - PL_0 - 10\gamma log_{10}(d) \tag{4}$$

Having 15 gateways and 20 nodes in an area spanning several kms, it made no sense trying to create a single regression that would possibly average the different environment losses.

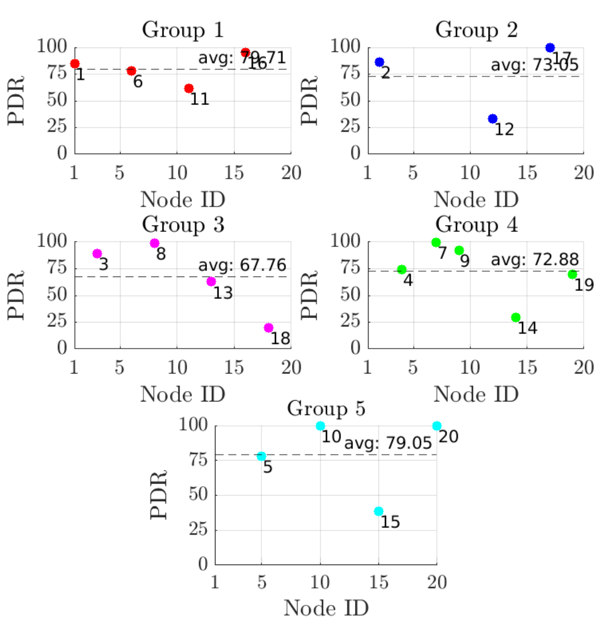


Fig. 3: PDR of devices per the different Tx Groups.

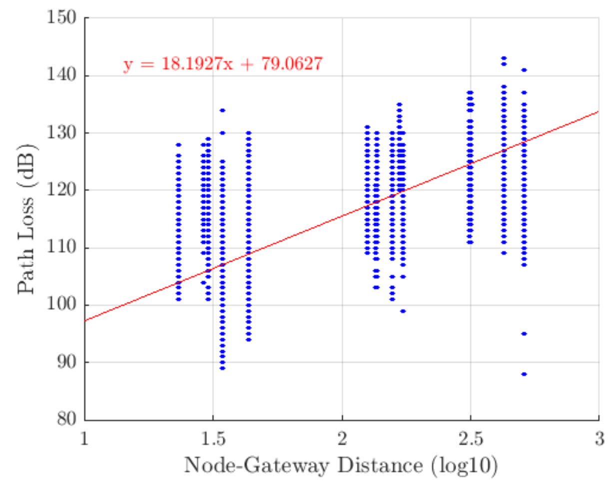


Fig. 4: Linear fit for the Path Loss recorded at Gateway 11.

Instead, the focus was on the packets received at Gateway 11, as it was the one that contributed the most to the network.

The results of the regression, along with the scatter plot of path loss against the node-gateway distance, are shown in Fig. 4.

Linked to the main, distance-based propagation loss model, we also included fast fading, provided by the NS-3 Nakagami-m model, approximating Rayleigh fading, as well as slow fading and building-related losses, provided and documented by the authors of the LoRaWAN module [3]. A combination of the different distance-based propagation models, plus effects of buildings, fast fading, and slow fading are tested to find which would best approximate our physical deployment.

B. Results and comparison

The various propagation loss chains were tested to find the most effective and conductive to a proper comparison. This was done by taking the average absolute difference between the PDR of the deployment and the PDR found by performing the different simulations. Table III shows the results for a number of combinations.

Traditional path loss schemes are only about 35% accurate.

The two closest solutions are a loss chain including the Friis free space model and the effects of buildings, fast fading and slow fading, and one including the Log-Distance model with the regression coefficients but no buildings effects. Unsurprisingly, the Log-Distance regression approximates the deployment the best, since we are gathering the fitting coefficients for γ and PL_0 (eq. 3) directly from the empirically obtained data. Buildings are not taken into account because they are already "included" in the regressed data, while Rayleigh fading as well as correlated slow fading are added to approximate the deviation from the regressed result. Where a regression is not feasible, as well as more generalised cases, using a loss chain with Friis, fast and slow fading, and building loss performed the best in this case.

From Fig. 5 it emerges that generally, the simulated and empirical PDR for the nodes fall within 20% of each other.

TABLE III: Some of the PDR differences between simulation and deployment achieved using different Path Loss chains.

Main propagation loss	Building related loss	Fast fading	Slow fading	Average PDR difference (%)
No path loss (perfect network)	X	X	X	25.14
Log-Distance with γ =3.76 and PL_0 = 7.7 at 1 metre	✓	✓	✓	34.72
Log-Distance with γ =3.92 and PL_0 = 11.52 at 1 metre	✓	✓	✓	41.15
Log-Distance with γ =3.92 and PL_0 = 11.52 at 1 metre	✓	✓	X	44.77
Log-Distance with regression coefficients, $\gamma = 1.819$ and $PL_0 = 79.063$ at 1 metre	✓	✓	✓	49.46
Log-Distance with regression coefficients, $\gamma = 1.819$ and $PL_0 = 79.063$ at 1 metre	X	✓	✓	21.3
Log-Distance with regression coefficients, $\gamma = 1.819$ and $PL_0 = 79.063$ at 1 metre	X	X	X	37.13
Okumura-Hata	✓	✓	✓	40.58
Friis	✓	✓	✓	24.25

However, when using the regression coefficients, the nodes that are placed in the more "challenging" environments will vastly differ in performance. Device 18 is the device in the harshest environment, being placed in a basement room, without windows, among electronic manufacturing equipment. The PDR obtained in the deployment reflects this, making it the device with the lowest recorded PDR. As there is no way to inform the simulation of the specifics of the room it is placed in, it is treated as others which are at a similar distance. The result is that its simulated PDR is about 60% better. The opposite is also true, as seen for devices 4, 7, 9 and 16. For these devices the simulation yields a worse PDR than what was recorded experimentally. This is also due to the topology of the rooms they were placed in, and it is a necessary consequence of obtaining the path loss coefficients through regression over nodes in different environments, in Fig. 4. For these four nodes, the regression fit lands above their Path Loss average, hence creating a worse performance in simulation than it should.

These discrepancies are also the case using building losses models. Without improving the existing NS-3 models and perfectly describing each building, the simulation is unaware of the exact layout, content, and precise position of each node in the room. Additionally, the existing building models for NS-3 that were investigated for this study are not concerned with the actual travel path of a signal, but only with checking if receiver and transmitter are indoor, outdoor, in the same building, or different buildings. This already is beneficial for standard, general models. Adding it to the regression however makes this path loss chain the worst performing one, as shown in Table III, with an average PDR difference of 49.46%.

Finally, a fundamental assumption made in the simulation is in fact that all gateways stayed active and operating throughout the whole length of the experiment. Other than speculation based on the timing of the reception of packets, this is not known, and the possible cause of further discrepancies.

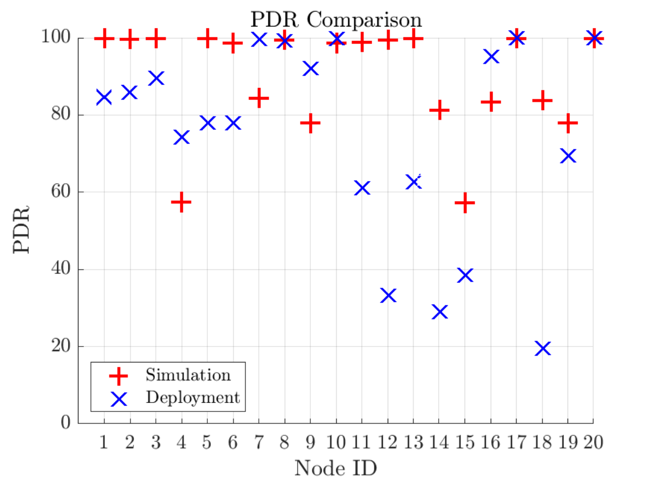


Fig. 5: Comparison between empirical and simulated PDR.

VI. Conclusion

As part of this study, a 20 devices LoRaWAN network was deployed in an urban environment and then simulated in software to appreciate the differences between the two methods. The 20 devices are split into 5 behaviour groups, to further diversify and characterise the nodes, and performance is compared in terms of simulated and empirical PDR. Results show that coupled with standard models for the additional losses such as the building effects, fast and slow fading, traditional path loss schemes are only partially accurate in recreating the losses experienced by our urban deployment. By using the empirically obtained RSSI data from the deployment itself, this difference in PDR can be cut down to roughly 20%. In our opinion, what we have presented in this work is the closest current technology allows simulations to get to a real life deployment, without spending an excessive amount of resources into modelling and recreating every minor detail. In future work, a more controlled environment could be used. Ideally this would include perfect information on all devices operating in the network proximity, their active periods and their exact positioning. A building loss model which takes into account the travel path of each signal, computing the approximate number of walls it must travel through, could also benefit this comparison. Excluding the ADR operation from the analysis would also prove beneficial. While strictly necessary for larger networks, in this case it caused more issues than it solved. This is possibly due to the "black box" approach of the ADR implementation on the network server, where no formal requirement is in place. Finally, aside from using dedicated energy-consumption models, devices should at least send their battery status as metadata to make better predictions on their total lifespan.

References

[1] Q. Liu, Y. Mu, J. Zhao, J. Feng, and B. Wang, "Characterizing Packet Loss in City-Scale LoRaWAN Deployment: Analysis and Implications," *IFIP Networking 2020 Conference and Workshops, Networking 2020*, pp. 704–712, 2020.

[2] R. Yasmin, J. Petajajarvi, K. Mikhaylov, and A. Pouttu, "Large and Dense LoRaWAN Deployment to Monitor Real Estate Conditions and Utilization Rate," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, vol. 2018-Septe, 2018.

[3] D. Magrin, M. Centenaro, and L. Vangelista, "Performance evaluation of LoRa networks in a smart city scenario," *IEEE International Conference on Communications*, 2017.

[4] O. Georgiou, C. Psomas, C. Skouroumounis, and I. Krikidis, "Optimal Non-Uniform Deployments of LoRa Networks," *IEEE Wireless Communications Letters*, vol. 2337, no. c, pp. 1–1, 2020.

[5] A. Mahmood, E. Sisinni, L. Guntupalli, R. Rondon, S. A. Hassan, and M. Gidlund, "Scalability Analysis of a LoRa Network under Imperfect Orthogonality," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1425–1436, 2019.

[6] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, and T. Watteyne, "Understanding the limits of lorawan," *IEEE Communications Magazine*, vol. 55, pp. 34–40, Sep. 2017.

[7] M. Cattani, C. A. Boano, and K. Römer, "An Experimental Evaluation of the Reliability of LoRa Long-Range Low-Power Wireless Communication," *Journal of Sensor and Actuator Networks*, vol. 6, no. 2, p. 7, 2017.

[8] M. Sharif and A. Sadeghi-Niaraki, "Ubiquitous sensor network simulation and emulation environments: A survey," *Journal of Network and Computer Applications*, vol. 93, no. May, pp. 150–181, 2017.

[9] G. Fortino, C. Savaglio, G. Spezzano, and M. Zhou, "Internet of Things as System of Systems: A Review of Methodologies, Frameworks, Platforms, and Tools," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 223–236, 2021.

[10] R. Muppala, A. Navnit, S. Poondla, and A. M. Hussain, "Investigation of Indoor LoRaWAN Signal Propagation for Real-World Applications," *2021 6th International Conference for Convergence in Technology, I2CT 2021*, pp. 17–21, 2021.

[11] M. Ohta, K. Adachi, N. Aihara, O. Takyu, T. Fujii, and M. Taromaru, "Measurement Experiments on 920 MHz Band for Spectrum Sharing with LoRaWAN," *IEEE Vehicular Technology Conference*, vol. 2018-Augus, 2018.

[12] J. D. F. Silveira, A. F. da S. Veloso, J. V. dos Reis Junior, A. C. B. Soares, and R. A. L. Rabelo, "A new low-cost lorawan power switch for smart farm applications," pp. 3330–3335, 1 2022.

[13] A. M. Yousuf, E. M. Rochester, and M. Ghaderi, "A low-cost lorawan testbed for iot: Implementation and measurements," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, pp. 361–366, Feb 2018.

[14] K. Mikhaylov, M. Stusek, P. Masek, R. Fujdiak, R. Mozny, S. Andreev, and J. Hosek, "On the Performance of Multi-Gateway LoRaWAN Deployments: An Experimental Study," *IEEE Wireless Communications and Networking Conference, WCNC*, vol. 2020-May, 2020.

[15] The Things Network, "The Things Uno."

[16] Semtech Corporation, "LoRaWAN – simple rate adaptation recommended algorithm," pp. 1–8, 2016.

[17] J. Petäjäjärvi, K. Mikhaylov, M. Hämäläinen, and J. Iinatti, "Evaluation of lora lpwan technology for remote health and wellbeing monitoring," in *2016 10th International Symposium on Medical Information and Communication Technology (ISMICT)*, pp. 1–5, March 2016.

[18] J. M. Marais, A. M. Abu-Mahfouz, and G. P. Hancke, "A Review of LoRaWAN Simulators: Design Requirements and Limitations," *Proceedings - 2019 International Multidisciplinary Information Technology and Engineering Conference, IMITEC 2019*, pp. 1–6, 2019.

[19] E. Harinda, S. Hosseinzadeh, H. Larijani, and R. M. Gibson, "Comparative Performance Analysis of Empirical Propagation Models for LoRaWAN 868MHz in an Urban Scenario," *IEEE 5th World Forum on Internet of Things, WF-IoT 2019 - Conference Proceedings*, pp. 154–159, 2019.

[20] O. Georgiou and U. Raza, "Low Power Wide Area Network Analysis: Can LoRa Scale?," *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 162–165, 2017.

[21] C. Caillouet, M. Heusse, and F. Rousseau, "Optimal SF allocation in LoRaWAN considering physical capture and imperfect orthogonality," *2019 IEEE Global Communications Conference, GLOBECOM 2019 - Proceedings*, 2019.

[22] C. Jiang, Y. Yang, X. Chen, J. Liao, W. Song, and X. Zhang, "A New-Dynamic Adaptive Data Rate Algorithm of LoRaWAN in Harsh Environment," *IEEE Internet of Things Journal*, vol. 4662, no. c, pp. 1–1, 2021.

[23] J. Petajajarvi, K. Mikhaylov, A. Roivainen, T. Hanninen, and M. Pettissalo, "On the coverage of lpwans: range evaluation and channel attenuation model for lora technology," in *2015 14th International Conference on ITS Telecommunications (ITST)*, pp. 55–59, Dec 2015.

[24] P. Jörke, S. Böcker, F. Liedmann, and C. Wietfeld, "Urban channel models for smart city IoT-networks based on empirical measurements of LoRa-lmks at 433 and 868 MHz," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, vol. 2017-October, no. Cn I, pp. 1–6, 2018.

Bruno Citoni received his BEng degree in Electronic Engineering with Music Technology Systems from the University of York, UK in 2016, and his MSc degree in Electronic and Electrical Engineering from the University of Glasgow, in 2017. He is currently working toward a PhD in the James Watt School of Engineering at the University of Glasgow, researching IoT, specifically LoRaWAN and its application for smart agriculture and smart cities.

Shuja Ansari [M'15, SM'20] received the M.Sc. degree (distinction) in Telecommunications Engineering in 2015, and the Ph.D. degree in Engineering in 2019 from Glasgow Caledonian University (GCU), UK. He is currently a Lecturer with the Glasgow College UESTC James Watt School of Engineering at the University of Glasgow (UofG) and Wave-1 Urban 5G use case implementation lead at Glasgow 5G Testbed (G5G) funded by the Scotland 5G Centre.

Qammer Hussain Abbasi (SM'16) received the Ph.D. degree in electronic and electrical engineering from the Queen Mary University of London (QMUL), U.K., in 2012. He is currently a Lecturer (Assistant Professor) with the School of Engineering, University of Glasgow, U.K. He has contributed to over 250 leading international technical journal and peer reviewed conference papers, and eight books.

Muhammad Ali Imran (M'03–SM'12) received the the Ph.D. degree from Imperial College London, U.K. in 2007. He is the Dean of Glasgow College, UESTC, and a Professor of Communication Systems with the School of Engineering, University of Glasgow. He is an Affiliate Professor with the University of Oklahoma, USA, and a Visiting Professor with the 5G Innovation Centre. He has been awarded 15 patents and (co)authored over 400 journal and conference publications

Sajjad Hussain [SM'17] is an Associate Professor (Senior Lecturer) in the James Watt School of Engineering, University of Glasgow, Scotland. His research interest include Industrial IoT, Energy-efficient solutions for 5G and beyond, self-organizing networks, Cognitive radio and cognitive radio networks, Smart grid communication and control, and Machine learning for wireless communications.