

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

A Configurable Mathematical Model for Single-Gateway LoRaWAN Performance Analysis

Davide Magrin, *Member, IEEE*, Martina Capuzzo, *Student Member, IEEE*,
Andrea Zanella, *Senior Member, IEEE*, and Michele Zorzi, *Fellow, IEEE*

Abstract

LoRaWAN is a Low Power Wide Area Network technology featuring long transmission ranges and a simple MAC layer, which can support sensor data collection, control applications and reliable services thanks to the flexibility offered by a large set of configurable system parameters. However, the impact of such parameters settings on the system's performance is often difficult to predict, depending on several factors. To ease this task, in this paper, we provide a mathematical model to estimate the performance of a LoRaWAN gateway serving a set of devices that may or may not employ confirmed traffic. The model features a set of parameters that can be adjusted to investigate different gateway and end-device configurations, making it possible to carry out a systematic analysis of various trade-offs. The results given by the proposed model are validated through realistic ns-3 simulations that confirm the ability of the model to predict the system performance with high accuracy, and assess the impact of the assumptions made in the model for tractability.

I. INTRODUCTION

The Internet of Things (IoT) makes it possible to remotely monitor and control a wide set of heterogeneous objects through an Internet connection. This paradigm foresees multiple applications in a large variety of scenarios: from fleet tracking and process monitoring in industrial scenarios to smarter garbage collection and intelligent light control in cities; from

The authors are with the Dept. of Information Engineering (DEI), University of Padova, Via Gradenigo 6/b, 35131 Padova, Italy. Profs. A. Zanella and M. Zorzi are also with the Human Inspired Technology (HIT) center of the University of Padova and with the Italian National Inter-University Consortium for Telecommunication (CNIT). Email: {magrinda, capuzzom, zanella, zorzi}@dei.unipd.it

monitoring of soil moisture in agriculture to home temperature control and personal health monitoring [1], [2], [3], [4]. Also, the IoT paradigm can be applied to surveillance-related applications [5], as event detectors and alarms [6].

The presence of several use cases spawned an ample market, and encouraged the development of multiple technologies meeting the need for low-cost ubiquitous connectivity. A large part of IoT nodes will consist in sensors that generate sporadic traffic, without strict constraints in terms of latency and throughput. This calls for new wireless solutions able to support a massive number of devices, with an affordable cost for both user equipment and network infrastructure. Therefore, high energy efficiency, extended coverage, and infrastructure simplicity are aspects of primary importance.

Such requirements motivated the creation of a new family of wireless technologies collectively called Low Power Wide Area Networks (LPWANs), characterized by long coverage range and low power consumption. A prominent LPWAN technology is LoRaWAN, which claims up to 10 years of battery lifetime for devices, and a transmission range between 1.5 km in urban scenarios and 30 km in rural areas [7].

Since the deployment of a dense IoT network is expensive and time consuming, performance assessments using simulations and mathematical models become essential to gauge the effect of network parameters and estimate the performance at a reduced cost. In this work, we propose an analytical model of the performance of a LoRaWAN network, accounting for the most relevant features of the LoRaWAN standard. This model considers network-layer performance, assuming perfect orthogonality between signals modulated with different Spreading Factors (SFs). However, compared to previous models in the literature (discussed in Sec. III), our model takes into account a wider range of aspects of the PHY and MAC layers, such as the possibility of transmitting multiple times both confirmed and unconfirmed packets, the limitations on the channel occupancy time imposed by the different national regulations, the interference produced by multiple overlapping transmissions, the capture effect, and the limited number of demodulators available at the Gateway (GW). Furthermore, the model formulation offers great flexibility in setting some system parameters, thus making it possible to analyze the system performance under different conditions and to shed light on possible trade-offs. We consider as performance metrics the packet success probability, average delays, and fairness, from which it is possible to derive other measures of interest, such as energy consumption, system's reliability and the achievable Quality of Service (QoS) in multiple scenario. The proposed model is validated by

comparing the results with those obtained through detailed ns-3 simulations. The analysis shows how the model can be used to maximize different performance metrics, proving a very powerful and convenient tool to determine the best network configuration.

The rest of this work is structured as follows. To make the paper self-contained, in Sec. II we present the main features of the LoRaWAN standard, while in Sec. III we give an overview of the current state of the art in the performance modeling of this technology. Sec. IV introduces the proposed model and describes how some of its parameters can be tuned to explore different behaviors of the network, while Sec. V briefly describes the simulation framework used for validation. Sec. VI, then, compares the output of the analytical and simulation models, also showing how they can be used to provide different insights of the network behavior. Finally, Sec. VII draws the conclusions and discusses possible future developments.

II. TECHNOLOGY OVERVIEW

This section describes the key LoRaWAN features, dwelling upon the elements and properties that have a significant impact on the system-level performance, which will then be considered in the model formulation.

A. The LoRa modulation

LoRa is a modulation technique based on Chirp Spread Spectrum (CSS), patented by Semtech. Bitrate and coverage range depend on the Spreading Factor (SF) parameter that can vary from 7 to 12. Lower SF values achieve higher data rates and shorter transmission times, but require higher signal powers at the receiver for correct decoding, which implies shorter coverage ranges. On the other hand, signals transmitted using higher SF values are more robust to channel impairments and can thus achieve longer transmission distances, at the price of an increased transmission time due to their lower data rates. Furthermore, signals modulated with different SFs are almost orthogonal: even if overlapping in time and frequency, two or more signals transmitted with different SFs can be simultaneously decoded, provided that their received powers satisfy some conditions [8].

When multiple packets transmitted with the same SF overlap in time and frequency, instead, they may generate destructive mutual interference, disrupting each other's reception and resulting in what is called a *packets collision* event. However, if one signal is significantly stronger than

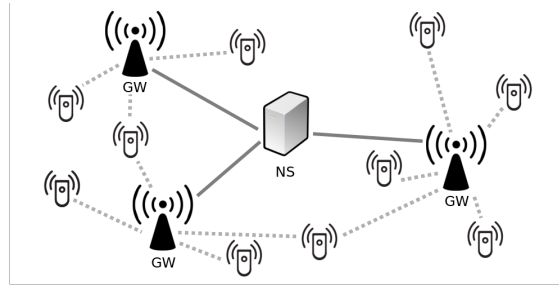


Figure 1: LoRaWAN network infrastructure. Dotted lines represent LoRa links, while solid lines represent IP connections.

the others, by a power margin greater than the so-called “co-channel rejection parameter” CR_{dB} , then it can be received correctly despite the interference, giving rise to a *capture* phenomenon.

The value of CR_{dB} has been estimated to be around 6 dB in [9]. In order to take advantage of these features, the SX1301 LoRa PHY chipset, typically employed in GWs [10], provides 8 parallel demodulation chains, which allow the chip to demodulate up to 8 different signals simultaneously, irrespective of their SFs and frequency. We also remark that the GWs do not support full-duplex transmission and reception: in order to send a Downlink (DL) packet they have to interrupt any ongoing reception, regardless of the frequency channels in which transmission and reception occur.

B. The LoRaWAN standard

The LoRaWAN standard [11] defines a star-of-stars topology, as represented in Fig. 1, with three kinds of devices: the *Network Server (NS)*, which is the central network controller and can be located anywhere in the Internet; the *End Devices (EDs)*, peripheral nodes (usually sensors or actuators) that transmit using the LoRa modulation; and the *Gateways (GWs)*, relay nodes that collect messages from the EDs through the LoRa interface and forward them to the NS using a reliable IP connection, and *vice versa*.

The standard also defines three classes of EDs, which differ for the time they spend in reception mode. This article considers *Class A* devices, which have the strictest requirements in terms of energy consumption. In order to save battery, these devices stay in sleep mode most of the time, opening two reception windows only 1 and 2 seconds after the end of an Uplink (UL) packet transmission. Fig. 2 depicts the different operational phases of a Class-A device when transmitting a UL packet and, then, receiving a DL packet in the second receive window.

The EDs have the possibility of transmitting *unconfirmed* or *confirmed* packets. In the latter case, an Acknowledgment (ACK) is expected in one of the two reception opportunities after

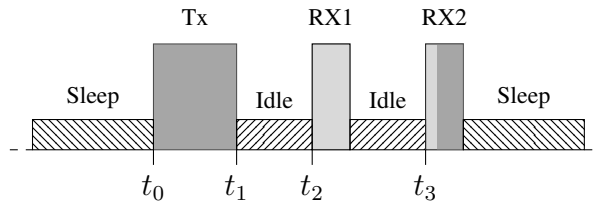


Figure 2: Example of operational phases for a Class-A ED. The device, initially in sleep mode, wakes up to transmit from time t_0 to t_1 . Then, the node remains in the idle state for 1 second, and at time $t_2 = t_1 + 1$ s opens the first receive window (RX1). If no packets are received, the device remains idle until the second receive window (RX2) is opened at time $t_3 = t_1 + 2$ s.

the transmission to confirm the correct reception of the packet by the NS.¹ If the ACK is not received in either of the two reception windows, a re-transmission can be performed at least RETRANSMIT_TIMEOUT seconds after the end of the second receive window. The standard recommends to randomly pick the value for RETRANSMIT_TIMEOUT uniformly between 1 and 3 seconds [12]. If the ACK is not received, the same confirmed message can be re-transmitted up to $m - 1$ times, after which the packet is dropped. The value of m can be configured by the NS.² To increase the robustness of unconfirmed transmissions, instead, the ED can transmit each packet h times. Once again, the value of h can be set by the NS. It is worth noting that the reception windows are opened after every UL transmission, irrespective of whether or not an ACK is expected, in order to give the NS the opportunity to send a DL packet to the EDs, if needed. The UL messages transmitted by an ED are collected by all the GWs in the coverage range of the transmitter, and forwarded to the NS. If the ED requires a reply, the NS can pick any of these GWs to transmit the DL message.

LoRaWAN operates in the ISM unlicensed spectrum, the use of which is subject to national regulations that define the maximum transmit power, and the Duty Cycle (DC) limit, i.e., the maximum percentage of time a node can actively transmit on a certain frequency band. The frequency bands, power and DC restrictions that apply to different regions are reported in the standard [12]. In particular, Table I shows the configuration mandated for the European region, which entails three bidirectional channels and a fourth channel reserved to DL transmissions only. The 868.1, 868.3, 868.5 MHz channels belong to the same regulatory sub-band (Sub-Band

¹Although in this paper we focus on ACK transmissions, the model and the analysis equally hold for any DL packet returned by the NS to the ED after the reception of a UL packet by the NS.

²This behavior holds for the LoRaWAN 1.1 standard [11], [12]; other versions of the standard may differ.

1 (SB1)), and have to share a DC limitation of 1%, while the channel reserved for DL, located in the 869 MHz sub-band (Sub-Band 2 (SB2)), can benefit from a more lenient DC of 10% and a higher transmission power.

Sub-band	Frequency [MHz]	Use	Duty Cycle
SB1	868.1	UL/DL	1%, shared
	868.3	UL/DL	1%, shared
	868.5	UL/DL	1%, shared
SB2	869.525	DL	10%, dedicated

Table I: Available LoRaWAN channels in the two sub-bands.

The SF used for a device's transmission is configured by the device itself or set by the NS according to some network management policies. By default, EDs open the first receive window (RX1) on the same frequency channel of the UL transmission, and expect a signal modulated with the same SF. The second receive window (RX2), instead, is opened on the 869.525 MHz channel and the incoming signal is assumed to use SF 12, to maximize the coverage rate. The standard allows the NS to modify this pre-defined configuration by communicating the new settings to the ED through appropriate Medium Access Control (MAC) commands, allowing for the use of any SF in the second window.

C. Packet life cycle

Messages transmitted by EDs to the GW are subject to multiple causes of losses:

- *Interference*: packets sent in the same frequency channel and with the same SF collide. A transmission can survive a collision event if its received power is sufficiently higher than that of the other overlapping signals (capture effect).
- *GW already in transmission*: the GW can not lock on a UL packet while performing a DL transmission.
- *GW starting a transmission*: an ongoing packet reception may be interrupted if the GW needs to send a DL packet.
- *No available demodulation chains at the GW*: all demodulators are already busy decoding incoming signals.

Moreover, confirmed UL messages cause the NS to generate ACKs that need to be transmitted by the GW. Such DL transmissions may as well be impaired by a number of events:

- *Unavailability of receive windows*: this event occurs when all available GWs are prevented from transmitting in both the receive windows because of the DC constraint or other ongoing transmissions.
- *Interference*: DL packets transmitted in RX1 can collide with UL packets transmitted by other EDs in the same channel and with the same SF.

In this work, we provide a network model that accounts for all these events.

III. STATE OF THE ART IN LORAWAN MODELING

In the last years, mathematical modeling has been applied to assess the network performance with respect to various metrics. In [13], the authors address high-level questions about LoRaWAN's suitability for a range of smart city applications, from metering to video surveillance, by modeling the system as a superposition of different Aloha networks. They conclude that, even if the long coverage range of a single GW makes the infrastructure able to serve several devices, the network must be carefully dimensioned to meet the application requirements. The work presented in [14] is one of the first to address the issue of scalability, using stochastic geometry to model interference in a LoRaWAN network. However, the study considers scenarios with only UL traffic. In [15] instead, queueing theory is applied to model latency and throughput of an ED subject to DC constraints, again focusing on UL communication only. The authors of [16], [17] provide a model based on Poisson arrival processes which takes DL communications, re-transmissions and capture effect into account. However, the analysis holds only in limited-size networks, where nodes can employ any transmission rate and their received powers are similar. The authors of [18] consider the features of the technology at the Physical (PHY) layer, by focusing on the capture effect and imperfect orthogonality between SFs: after performing empirical measurements, they model these effects and derive the throughput achieved by the network for different cell configurations and number of GWs. In [19], the problem of network scalability is faced through mathematical modeling and Python-based simulations, taking into account also the capture effect, and evaluating the impact of SF allocation and power control. In all these works, however, the main focus is on the PHY layer, and downlink traffic and re-transmissions are not considered. Finally, the work presented in [20] proposes a model to calculate energy consumption and delay for reliable UL traffic in a LoRaWAN network. The

results for a limited number of devices are compared to real test-bed measurements and to the outcome of ns-3 simulations. The analysis, based on Markov-chain theory, neglects the DC constraints in the different sub-bands, and assumes that ACKs are always sent in one specific receive window (either RX1 or RX2). Markov chains are also proposed in [21] to characterize the performance of a LoRaWAN battery-less device; however, the study considers a single device, and the network analysis is left for future work.

The work presented in this paper is an extension of our previous conference paper [22], where we modeled a wide network with Poisson packet arrivals, considering the DC limitations and a set of network parameters. Here we revise the model by developing a novel approach to accurately consider the limited availability of reception chains at the GW, the peculiarities of the two receive windows, and the DC constraints. Additionally, we include packet re-transmissions and the capture effect. Compared to the state of the art, our model includes the ability to take into account the coexistence of unconfirmed and confirmed traffic and, at the same time, maintains the possibility of estimating the network behavior under several network configurations with minimal effort. The results obtained through this model are compared with those given by a state-of-the-art and open source LoRaWAN simulator, presented in [23], further attesting the accuracy of the proposed approach and exploring the impact of common assumptions. Finally, we also show some possible usages of the model to evaluate a wide variety of network configurations with limited effort.

IV. MODEL

The aim of the model proposed in this paper is to characterize the behavior of a LoRaWAN network with a single GW, which receives packets from a set of EDs and needs to reply in one of the two receive windows when an ED requires confirmation. The system performance is assessed in terms of packet success probability, following the approach used in [22] and extending it with a more accurate characterization of the GW behavior. This performance metric is proxy to other fundamental metrics, such as throughput and network capacity, which can be straightforwardly derived from it. The following sub-sections are structured as follows. The reference scenario, model assumptions, system parameters and their effects are described in Sec. IV-A, together with a brief presentation of the structure of the model and its underlying rationale; Sec. IV-B, then, describes some relevant quantities and parameters of the proposed model. We then delve into the analytical formulation by decoupling the analysis of the UL traffic (Sec. IV-C and IV-D) and DL

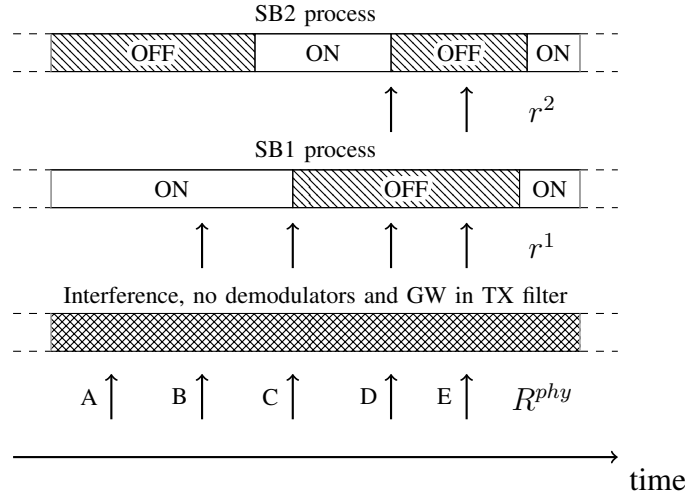


Figure 3: Representation of the model's packet filtering structure. R^{phy} is the rate of UL traffic (see (4)), while r^1 and r^2 represent the rate of ACKs sent in SB1 and SB2, respectively (see (16), (18)).

messages (Sec. IV-E), and derive the formulas for DL success probabilities in Sec. IV-F. Finally, Sec. IV-G, describes different performance metrics and their computation. Note that, because of the mutual dependency of some values, some terms may be described and introduced before the corresponding equation can be derived, in which case references are provided in the text.

A. Scenario and assumptions

We consider a scenario where the EDs are randomly and uniformly distributed around a single GW. Application-layer packets are generated according to a Poisson Process with aggregate packet generation rate λ [pck/s], and can be either confirmed or unconfirmed.

For tractability, we assume perfect orthogonality between different SFs, i.e., only packets employing the same SF can collide. In this case, one of the two packets can survive if its received power is sufficiently higher than that of the colliding packet (collisions with more than two packets happen with negligible probability and are not considered). While the orthogonality assumption has been shown to have an impact on the PHY-layer performance of UL only traffic [8], the results discussed in Sec. VI show that the effect is much more limited in the presence of confirmed traffic, where the performance is severely limited by other factors.

Fig. 3 shows the structure of the packet reception model, consisting in successive filtering of Poisson processes. At the base of the figure, arrows are used to represent the UL traffic generated by the EDs, including both new packet transmissions and re-transmissions of failed packets.

This process is assumed to be Poisson for tractability, ignoring the fact that re-transmissions of a certain packet are correlated in time because of DC limitations. An initial filtering of this process excludes some arrivals, modeling packet losses due to interference from other EDs, unavailability of GW demodulators, or ongoing DL transmissions from the GW. This yields a process with a reduced rate, which now represents the packets that are correctly received by the GW.

When the received UL message requires confirmation, an ACK must be sent by the GW during one of the two receive windows of the target ED. The ability of the GW to perform such a transmission is modeled through two independent alternating renewal processes, in which the system alternates between the ON and OFF states. The two processes represent the opportunity of sending the ACK in RX1 or RX2, respectively, which are opened on SB1 or SB2, i.e., on the shared or dedicated sub-band. If a confirmed packet finds a process in the ON state, it means that the GW will be able to send an ACK in that sub-band. In this case, the process will switch to the OFF state to model the unavailability of that sub-channel for a certain period of time following the ACK transmission, due to the DC restrictions.

Since the sub-bands are disjoint, we assume that the two processes are uncorrelated, neglecting the fact that the very packets that need to be served in SB2 are those that found SB1 in the OFF state. If the DL packet finds at least one of the two processes in the ON state, an ACK is sent. If the ACK is sent on SB1 (hence, using frequencies shared by UL and DL traffic), it can be destroyed by the interference created by other EDs. If the ACK is sent on SB2, instead, it is assumed to be always successful.

For the sake of clarity, the following list describes some examples of the life cycle of the packets in Fig. 3:

- (A) This packet is lost because of interference or GW transmission or unavailability of demodulators. Hence, it does not pass the first filter.
- (B) This is an unconfirmed UL packet, which is successfully received by the GW. It does not generate any ACK.
- (C) This is a confirmed packet successfully received by the GW. It generates an ACK, which finds the SB1 process in the ON state. The ACK is successfully sent, and the SB1 process switches to the OFF state.
- (D) This is another confirmed packet which is successfully received by the GW. Since the GW has just sent an ACK for packet (C), it cannot reply in SB1 due to DC constraints; SB2 is

however in the ON state, and the GW can thus reply to the ED, making the second process switch to the OFF state.

- (E) This is another confirmed packet, which gets a treatment similar to that of packet (D). However, since the GW has transmitted the ACK for packet (D) and is still under the DC constraints, it cannot reply to packet (E) in either of the two receive windows (both SB1 and SB2 processes are in the OFF state). The DL packet is hence discarded, and the ED will re-transmit the UL message at a later time.

B. Model Quantities

Our model offers some tunable parameters to increase its flexibility, enabling the evaluation of the network performance in various configurations with minimal effort. The model makes it possible to specify the following values:

- $\mathcal{SF} = \{7, \dots, 12\}$ indicates the set of all SFs.
- α : fraction of application-layer traffic requiring confirmation;
- p_i^u, p_i^c : fraction of devices generating unconfirmed and confirmed traffic with a specific SF $i \in \mathcal{SF}$, respectively. Note that $\sum_{i \in \mathcal{SF}} p_i^u = \sum_{i \in \mathcal{SF}} p_i^c = 1$;
- h : number of times an application-layer unconfirmed packet is transmitted;
- m : maximum number of transmission attempts for confirmed packets;
- δ_{SB1} and δ_{SB2} : ratio between silent time and transmission time in SBk , corresponding to the DC constraint. For instance, in Europe, we have $\delta_{SB1} = 99$ and $\delta_{SB2} = 9$ corresponding to a DC of 1% in SB1 and 10% in SB2. In general, when $\delta_{SBk} > 0$ the DC constraint applies to all devices transmitting in subchannel SBk . Instead, the setting $\delta_{SBk} = 0$ corresponds to a DC constraint of 100%, which means that there is no limitation on the transmission time³;
- τ_1 and τ_2 : prioritization flags. If $\tau_k = 0$, the GW prioritizes reception operations over transmission during the k -th receive window, with $k = 1, 2$. In this case, the GW will drop any DL message that needs to be transmitted while a UL reception is ongoing. Instead, if TX is prioritized ($\tau_k = 1$), the reception of any incoming packet will be interrupted in order to send the ACK;

³This setting is not allowed by current RF recommendations but is considered in this study to gain insights on the impact of DC limitations in the considered scenarios.

- C : number of UL frequency channels. Note that each UL channel can also be used for DL transmissions. Instead, the channel in SB2 is DL only;
- T_i^{ack2} : duration of the transmission of the ACK in RX2 when using SF i . (The standard requires the use of SF 12 in RX2 as a pre-configured setting, corresponding to T_{12}^{ack2} . Note that this default setting can be changed by the NS, and accordingly in our model.)
- T_i^{data} and T_i^{ack1} indicate the time durations of a data packet and of an ACK transmitted in SB1 with SF i , respectively. If ACKs transmitted in SB2 use SF12, irrespective of the SF employed in the UL transmission, then $T_i^{ack2} = T_{12}^{ack1}$, $\forall i \in \mathcal{SF}$.

In the formulas, the notation generally respects the following scheme. The probability is indicated with S or F if it corresponds to a “success” or “failure” event, respectively; if this rule does not apply, the probability is denoted simply as P . The superscript indicates the considered event, while the subscript the SF. For example, in (6), the symbol S_i^{INT} represents the probability of successfully surviving interference when using SF i . Different uses of the notation are specified in the text. The following sections provide a mathematical formulation for some relevant quantities in this model.

C. Uplink traffic rates

The assumption of perfect orthogonality between different SFs makes it possible to split the network traffic in different logical channels that do not interfere with each other. The traffic load for each SF i is split uniformly over the given C frequency channels (since EDs pick a random UL frequency for each transmission attempt). Thus, the traffic generated at the application layer by the EDs using confirmed and unconfirmed messages is, respectively, given by:

$$R_i^{c,app} = \frac{P_i^c \cdot \lambda}{C} \cdot \alpha, \quad (1)$$

$$R_i^{u,app} = \frac{P_i^u \cdot \lambda}{C} \cdot (1 - \alpha). \quad (2)$$

Since EDs using unconfirmed traffic will perform h transmissions of each application-layer packet, the PHY rate of these devices can be computed as $R_i^{u,phy} = R_i^{u,app} \cdot h$. For EDs transmitting confirmed messages, instead, the number of re-transmitted packets depends on the success of both the UL transmission and the corresponding ACK. We indicate as $P_{i,j}^{DL}$ the probability that a confirmed UL packet sent with SF i is successfully received and acknowledged at the j -th transmission attempt, which will be derived in (29). Therefore, we have that the rate of confirmed

packets transmitted at SF i , $R_i^{c,phy}$, is given by the product of the application-level rate, $R_i^{c,app}$, and the average number of times a confirmed packet is transmitted at the PHY layer.

$$R_i^{c,phy} = R_i^{c,app} \left[\sum_{j=1}^{m-1} j \cdot P_{i,j}^{DL} + m \left(1 - \sum_{j=1}^{m-1} P_{i,j}^{DL} \right) \right]. \quad (3)$$

The first summation in the square brackets of (3) takes into account transmissions that are successfully received before the m th attempt, while the second term considers the case when the packet is transmitted m times (irrespective of whether the last transmission is successful or not).

The total traffic for a single frequency channel and for SF i is therefore given by

$$R_i^{phy} = R_i^{u,phy} + R_i^{c,phy}. \quad (4)$$

In general, the distribution of the SFs for the transmitted packets at the PHY layer will differ from the native distribution of SFs among the devices, $\{p_i^u, p_i^c\}$, because of re-transmissions.

Thus, we define

$$d_i = \frac{R_i^{phy}}{\sum_j R_j^{phy}}, \quad (5)$$

as the ratio of PHY layer packets that are transmitted at SF $i \in \mathcal{SF}$.

D. PHY layer probabilities

A UL packet is successfully received by the GW if all the following conditions are met: (i) it does not overlap with another UL transmission using the same SF on the same frequency, or it overlaps with another UL packet, but the received power is sufficiently large to allow for correct decoding despite the interference (capture), (ii) it does not overlap with a GW DL transmission in any channel, and (iii) it finds an available demodulator. These conditions are represented by the first filter in Fig. 3.

Since packets are generated following a Poisson process, the probability of event (i) is given by two components. The first is the probability that there are no other arrivals during the $2T_i^{data}$ vulnerability period across the packet arrival instant. The second, considers a collision with one packet, and the fact that the receiver successfully captures the frame. For the UL, we consider the capture probability \mathbb{W}^{GW} as computed in [16]. Since these two events are disjoint, the probability

of surviving interference (event (i)) is given by the sum of the two components, which results in

$$S_i^{INT} = e^{-2T_i^{data}R_i^{phy}} + 2T_i^{data}R_i^{phy}e^{-2T_i^{data}R_i^{phy}} \cdot \mathbb{W}^{GW}, \quad (6)$$

where, in the right-most term, we computed the probability that either of the two colliding packets is captured (collision events with more than two packets are neglected).

To compute the probability of event (ii), we observe that a UL message is always lost when it arrives at the GW during the transmission of an ACK. Otherwise, the GW will start the reception of the UL message, which will take a time T_i^{data} . If reception on SB k is prioritized (i.e., $\tau_k = 0$), this process cannot be interrupted, and the UL message will be successfully delivered to the NS. Conversely, if $\tau_k = 1$, i.e., we prioritize transmission on SB k , the reception of the UL packet may be aborted at any time during the period T_i^{data} , in order to give priority on the ACK transmission. Therefore, the vulnerability period is given by the ACK transmission time T_s^{ackk} , to which we need to add the interval T_i^{data} only if $\tau_k = 1$. Denoting by b_s^k the probability that an ACK is transmitted on SB k with SF $s \in \mathcal{SF}$ (which will be derived later in (22)), the average vulnerability period is then given by $\bar{T}_k = \sum_{s \in \mathcal{SF}} b_s^k T_s^{ackk} + T_i^{data} \cdot \tau_k$. Now, according to the Poisson Arrivals See Time Averages (PASTA) property, the probability that a UL packet arrival falls in the vulnerability period of channel SB k , with $k = 1, 2$, can be expressed as

$$F_i^{TXk} = \frac{\sum_{s \in \mathcal{SF}} b_s^k T_s^{ackk} + T_i^{data} \cdot \tau_k}{E_{ON}^k + E_{OFF}^k}, \quad (7)$$

where the denominator is the mean renewal time of the SB k process, given by the sum of E_{ON}^k and E_{OFF}^k , i.e., the expected times the SB k process spends in the ON and OFF states during a renewal period (ON-OFF cycle), which will be computed in (21) and (23). Then, assuming (for ease of analysis) that events in SB1 and SB2 are independent, the probability that a UL packet is successfully received (event (ii)) is given by

$$S_i^{TX} = (1 - F_i^{TX1})(1 - F_i^{TX2}). \quad (8)$$

Next, we compute the probability of event (iii), i.e., that at least one demodulator out of 8 is available. Each demodulator chain is modeled through an alternating renewal process, where the demodulator can be in an ‘‘available’’ state A , when idle or in a ‘‘locked’’ state L , when occupied with the reception of another signal. We assume that the different demodulators are activated in succession: if all are available, an incoming signal will be received by the first demodulator; if the first demodulator is in the L state, the packet will be handled by the second demodulator,

and so on. Let E^L be the expected time a demodulator will be locked on a incoming signal. Since the occupation will last for the duration of UL LoRa packets at the PHY layer, we have:

$$E^L = \sum_{i \in \mathcal{SF}} d_i \cdot T_i^{data}. \quad (9)$$

The average time the first demodulator is in the A state, instead, is computed as the average inter-arrival time of UL packets, regardless of their SF and selected frequency:

$$E^{A,1} = \frac{1}{C \cdot \sum_{i \in \mathcal{SF}} R_i^{phy}}. \quad (10)$$

Then, the process of packets that require the second demodulator is filtered by the probability of finding the first demodulator occupied. Thus, the expected time the second demodulator is available is given by

$$E^{A,2} = \frac{E^{A,1}}{P^{L,1}} = \frac{1}{P^{L,1} \cdot C \cdot \sum_{i \in \mathcal{SF}} R_i^{phy}}, \quad (11)$$

where $P^{L,1}$ is the probability that the first modulator is in the L state (see (13)). With a similar reasoning, we compute the expected time for which the j -th demodulator is available as

$$E^{A,j} = \frac{E^{A,j-1}}{P^{L,j-1}} = \frac{1}{\prod_{\ell=1}^{j-1} P^{L,\ell} \cdot C \cdot \sum_{i \in \mathcal{SF}} R_i^{phy}}. \quad (12)$$

The probability $P^{L,\ell}$ of finding the ℓ -th demodulator in the L state, in turn, can be expressed as

$$P^{L,\ell} = \frac{E^L}{E^{A,\ell} + E^L}. \quad (13)$$

Then, a packet finds an available demodulator (event (iii)) with probability:

$$S^{demod} = 1 - \prod_{j=1}^8 P^{L,j}. \quad (14)$$

The overall UL packet success probability, considering events (i), (ii) and (iii) described above, is finally expressed as

$$S_i^{UL} = S_i^{INT} \cdot S_i^{TX} \cdot S^{demod}. \quad (15)$$

E. ACK transmission

Once a confirmed packet is correctly received by the GW, an ACK needs to be transmitted back to the ED. Eq. (15) gives the probability of successful packet reception at the GW. Therefore, the rate of ACK messages that the GW will try to send in SB1 is:

$$r_i^1 = R_i^{c,phy} \cdot S_i^{UL}. \quad (16)$$

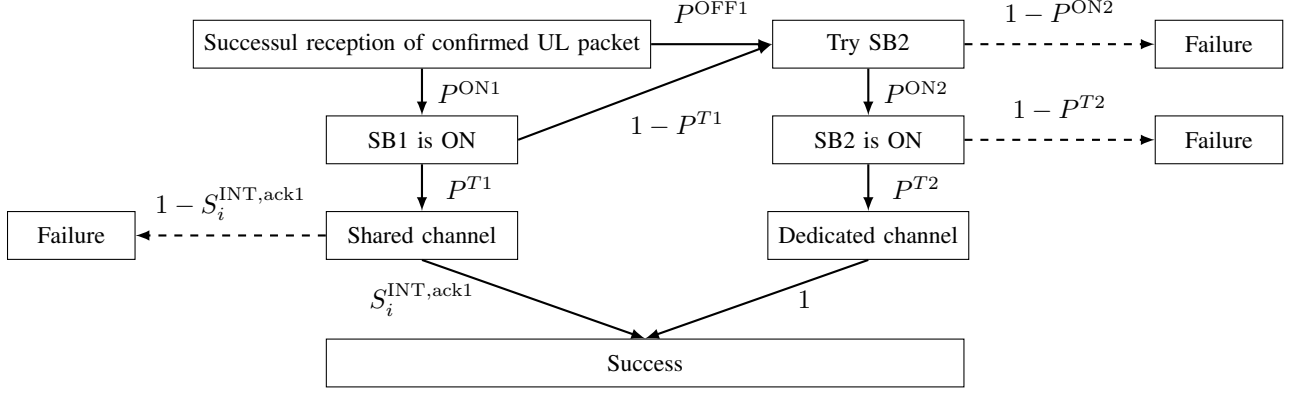


Figure 4: Diagram for successful ACK reception.

A visual representation of the possible ACK life cycles considered in the model is shown in Fig. 4. Labels refer to the probabilities of the different events, which we derive next. In general, an ACK is transmitted in SBk if both the following conditions hold: (i) $\tau_k = 1$ (TX is prioritized) or $\tau_k = 0$ and the GW is idle; (ii) SBk is available (i.e., not blocked by DC constraints). If either condition is not satisfied, the ACK is dropped.

Let T denote the event “the GW *may* transmit,” which depends on the TX/RX prioritization policy. If $\tau_k = 1$, the GW can transmit the DL packet whenever it needs to; otherwise, if $\tau_k = 0$, the GW can transmit in SBk only if no reception is ongoing. We denote by $P^{T,k}$ the probability of T , which can be computed as

$$P^{T,k} = \begin{cases} 1, & \text{if } \tau_k = 1; \\ e^{-\sum_{i \in \mathcal{SF}} C \cdot R_i^{phy} T_i^{data}}, & \text{if } \tau_k = 0; \end{cases} \quad (17)$$

where the second expression is the probability that no UL packet was generated in the last T_i^{data} seconds.

If $SB1$ is not available, the GW will try to process the ACK in $SB2$. Such packets form a process with rate

$$r_i^2 = r_i^1 [P^{OFF,1} + P^{ON,1}(1 - P^{T,1})], \quad (18)$$

where $P^{ON,1}$ and $P^{OFF,1}$ are the probabilities of finding $SB1$ in the ON and OFF state, respectively, and $(1 - P^{T,1})$ is the probability that the GW is not available for DL transmission. The

ON and OFF probabilities for the SB k process, with $k = 1, 2$, are given by

$$P^{ON,k} = \frac{E^{ON,k}}{E^{ON,k} + E^{OFF,k}}, \quad (19)$$

$$P^{OFF,k} = \frac{E^{OFF,k}}{E^{ON,k} + E^{OFF,k}}, \quad (20)$$

where $E^{ON,k}$ and $E^{OFF,k}$ are the mean sojourn times in ON and OFF states, respectively, which are computed as follows. By considering the arrival rate of successful UL packets in the k -th sub-band, we have:

$$E^{ONk} = \frac{1}{\sum_{i \in \mathcal{SF}} C \cdot r_i^k}. \quad (21)$$

Note that the switch from the ON to the OFF state will be caused by a packet sent in any of the C UL channels: therefore, we need to multiply the rates r_i^k of arrivals to SB k with SF i by the number of available channels.

In order to compute the expected duration of the OFF periods, we first need to derive the probability distribution b_i^k of the SFs used for ACK transmissions, which is given by

$$b_i^k = \frac{r_i^k}{\sum_{s \in \mathcal{SF}} r_s^k}. \quad (22)$$

In our model, the OFF period accounts for the time the GW is prevented from transmitting a new data packet, which includes the time to send the ACK using the given SF, plus the waiting time imposed by the DC limitations. We hence have

$$E^{OFF,1} = \sum_{s \in \mathcal{SF}} b_s^1 (T_s^{\text{ack}1} + \delta_{SB1} \cdot T_s^{\text{ack}1}), \quad (23)$$

$$E^{OFF,2} = \sum_{s \in \mathcal{SF}} b_s^2 (T_s^{\text{ack}2} + \delta_{SB2} \cdot T_s^{\text{ack}2}).$$

(Note that, by including the parameter δ_{SBk} as defined in Sec. IV-A, we can change the DC limitations in the k -th sub-band, thus making it possible to analyze its impact.)

Finally, we remark that DL packets sent by the GW in SB1 also have to avoid interference from other EDs. In the absence of collisions, the vulnerability period is given by the sum of two terms. The first term corresponds to the case of no UL transmissions starting while the DL packet is being sent ($T^{\text{ack}1}$); the second term represents the event where no UL transmissions started before the ACK is sent. Note that if $\tau_1 = 0$ the second term is not present, since in that case the ACK would not be generated at all. Furthermore, an ACK can survive an interfering packet sent by another ED in case of capture, which happens with probability \mathbb{W}^{ED} (equivalent

to the \mathbb{W}^{Mote} as derived in [16]). Therefore, the probability that the ACK does not collide with a UL packet in SB1, or is captured despite the collision, is equal to

$$S_i^{INT,ack1} = e^{-R_i^{phy}(T_i^{ack1} + \tau_1 \cdot T_i^{data})} + R_i^{phy}(T_i^{ack1} + T_i^{data}) \cdot e^{-R_i^{phy}(T_i^{ack1} + T_i^{data})} \cdot \mathbb{W}^{ED}. \quad (24)$$

For packets sent in SB2, instead, the reception is assumed to be always successful, since the 869.525 MHz channel is dedicated to DL communication and the GW only transmits one packet at a time (note that this assumption does not hold in the case of multiple GWs).

F. DL success probability

Given that a confirmed UL packet sent with SF i has been successfully received by the GW, the probability that the corresponding ACK is also successfully returned to the ED is expressed as

$$S_i^{DL} = S_i^{SB1} + S^{SB2}, \quad (25)$$

where S_i^{SB1} describes the probability of a successful ACK transmission in SB1 with SF i , while S^{SB2} accounts for the probability that SB1 is not available, and the ACK is successfully sent in SB2. These probabilities, in turn, can be expressed as follows:

$$S_i^{SB1} = P^{ON,1} \cdot P^{T,1} \cdot S_i^{INT,ack1}, \quad (26)$$

$$S^{SB2} = [P^{OFF,1} + P^{ON,1} \cdot (1 - P^{T,1})] \cdot P^{ON,2} \cdot P^{T,2}. \quad (27)$$

Fig. 4 can be used as a reference for the computation of this quantity.

Finally, we can compute the success probabilities over m transmissions. We recall that, for the sake of simplicity, we neglect the time correlation of packet re-transmissions due to DC constraints, (the impact of this approximation will be analyzed by simulation). We recall that $P_{i,j}^{UL}$ indicates the probability that a UL packet with SF i is successfully received at the GW at exactly the j -th transmission attempt, which can be computed as:

$$P_{i,j}^{UL} = S_i^{UL} (1 - S_i^{UL})^{j-1}. \quad (28)$$

Then, the ED successfully receives the ACK at exactly the j -th attempt if both the UL and the DL transmissions succeed. The probability $P_{i,j}^{DL}$ of this event is hence given by:

$$P_{i,j}^{DL} = [1 - (S_i^{UL} S_i^{DL})]^{j-1} \cdot (S_i^{UL} S_i^{DL}). \quad (29)$$

Once all intermediate quantities are computed, the model can be summarized by two inter-dependent equations:

$$\begin{cases} S^{UL} = f(S^{UL}, S^{DL}), \\ S^{DL} = g(S^{UL}, S^{DL}). \end{cases}$$

where $S^{UL} = [S_7^{UL}, \dots, S_{12}^{UL}]$ and $S^{DL} = [S_7^{DL}, \dots, S_{12}^{DL}]$, while $f()$ and $g()$ are implicit functions given by the chaining of the sequence of operations that yield (15) and (25), respectively.

This system admits a fixed-point solution, which can be found through fixed-point iteration. From a practical perspective, when initialized with the states $S^{UL} = S^{DL} = [1, 1, 1, 1, 1, 1]$, the iterative process has always reached convergence to the stable fixed point after a few iterations (order of few units) for all the parameter combinations considered in this work. The proof of the system's convergence is provided in [24]. An implementation of the model, allowing the interested readers to easily replicate the results shown in this paper, is publicly available at [25].

G. Performance metrics

To evaluate the system performance, we consider three classes of key performance indicators, namely: reliability, delay, and fairness metrics which are better detailed in the remainder of this section together with the methodology to determine their value using the proposed model. Once a set of parameters is fixed, the model can be solved and the performance metrics can be estimated starting from S^{UL} and S^{DL} . Conversely, it is possible to employ the model to optimize a given performance metric, finding the parameter setting that maximizes it, as shown in Sec. VI.

1) *Reliability Metrics*: We consider three Packet Delivery Rate (PDR) indexes, namely:

- *Unconfirmed Uplink PDR (UU)*: fraction of (application-layer) unconfirmed packets that are successfully received by the GW;
- *Confirmed Uplink PDR (CU)*: fraction of (application-layer) confirmed packets that are successfully received by the GW, irrespective of whether or not the corresponding ACK is successfully returned to the ED;
- *Confirmed Downlink PDR (CD)*: fraction of (application-layer) confirmed packets that are successfully acknowledged by the NS.

Clearly, $CD \leq CU$, since a packet needs to be successfully received by the GW in order to be acknowledged. Note that the CU metric captures the performance of applications for which it is important to deliver packets to the NS and ACKs are only used to stop re-transmissions

(and thus avoid a useless increase in traffic), while CD is more interesting for applications that require the EDs to get explicit feedback from the NS, for instance containing control information addressed to the ED.

We obtain the UU and CU values by averaging the UL success probability (UU_i and CU_i for unconfirmed and confirmed packets, respectively) for each SF i over the SF distribution, i.e.,

$$UU = \sum_{i \in \mathcal{SF}} (p_i^u \cdot UU_i) = \sum_{i \in \mathcal{SF}} \left(p_i^u \cdot \sum_{j=1}^h P_{i,j}^{UL} \right), \quad (30)$$

$$CU = \sum_{i \in \mathcal{SF}} (p_i^c \cdot CU_i) = \sum_{i \in \mathcal{SF}} \left(p_i^c \cdot \sum_{j=1}^m P_{i,j}^{UL} \right). \quad (31)$$

Similarly, CD is computed as the probability of success for a confirmed packet within the available re-transmission attempts

$$CD = \sum_{i \in \mathcal{SF}} \left(p_i^c \cdot \sum_{j=1}^m P_{i,j}^{DL} \right). \quad (32)$$

2) *Delay Metrics*: We define two delay metrics, considering confirmed traffic only: Δ^{UL} measures the time from the first transmission attempt to the successful delivery to the GW of an UL confirmed packet, while Δ^{DL} accounts for the time from the first transmission of a confirmed packet to the successful reception of the corresponding reply. Delays are computed for successful packets only, and the propagation delay is assumed to be negligible. To compute these metrics with our model, we assume the RETRANSMIT_TIMEOUT value to be a uniformly distributed random variable with mean μ , and consider that EDs employ the shared sub-band with δ_{SB1} DC limitations. Therefore, the average time between two transmissions of the same MAC-layer packet by a device is given by:

$$\gamma_i = (\delta_{SB1} + 1) \cdot T_i^{data} + \mu. \quad (33)$$

The average delay from the successful reception of a packet at the GW to the transmission of the ACK is given by:

$$\phi_i = S_i^{SB1} \cdot (1 + T_i^{ack1}) + S_i^{SB2} \cdot (2 + T_i^{ack2}), \quad (34)$$

where we take into account that the ACK will be served in SB1 (opened after 1 second) with probability S_i^{SB1} , and in SB2 (opened after 2 seconds) with probability S_i^{SB2} .

If a packet is re-transmitted m times, each re-transmission j is associated with a certain UL success probability $P_{i,j}^{\text{UL}}$. The average delay at each SF $i \in \mathcal{SF}$ can be computed as:

$$\Delta^{\text{UL}} = \sum_{i \in \mathcal{SF}} p_i^c \cdot \left(\sum_{j=1}^m \bar{P}_{i,j}^{\text{UL}} (T_i^{\text{data}} + (j-1) \cdot \gamma_i) \right), \quad (35)$$

where we define $\bar{P}_{i,j}^{\text{UL}} = P_{i,j}^{\text{UL}} / \sum_j P_{i,j}^{\text{UL}}$ to obtain the distribution of successful UL packet transmissions.

Similarly, we can compute the average ACK delay:

$$\Delta^{\text{DL}} = \sum_{i \in \mathcal{SF}} p_i^c \cdot \left(\sum_{j=1}^m \bar{P}_{i,j}^{\text{DL}} (T_i^{\text{data}} + (j-1) \cdot \gamma_i + j \cdot \phi_i) \right), \quad (36)$$

where, in addition to the inter-transmission time between two packets, we also account for the time to perform the ACK transmission.

3) *Fairness*: Finally, we consider the fairness of the system in different scenarios. Indeed, EDs employing confirmed traffic or higher SFs will use more system resources (e.g., channel occupancy), possibly affecting the application performance of devices that employ different settings. To this aim, we use Jain's fairness index, defined as

$$J(\mathbf{x}) = \frac{\left(\sum_{i=1}^n x_i \right)^2}{n \cdot \sum_{i=1}^n x_i^2}, \quad (37)$$

where n is the total number of user categories, each with throughput x_i . Note that $1/n \leq J(\mathbf{x}) \leq 1$, and the system is perfectly fair if $J(\mathbf{x}) = 1$. In particular, in the following section, we will consider the fairness among devices employing different SFs. Furthermore, since all the devices have equal packet generation rate, and transmit packets with the same length, instead of the throughput we can simply consider the UL success probability, i.e., UU for nodes employing unconfirmed traffic and CU for devices transmitting confirmed messages. Therefore, the fairness is computed by taking $\mathbf{x} = [\mathbf{x}^u, \mathbf{x}^c]$, where the elements correspond to $x_i^u = UU_i$, and $x_i^c = CU_i$, as defined in (30), (31).

V. NETWORK SIMULATIONS

In order to validate our model, we compared the performance estimates obtained from the model with those observed in more realistic simulations, in which most of the simplifying assumptions of the model are removed.

This section describes how we employ the LoRaWAN ns-3 module described in [26] to perform such a validation. To be noted that the more accurate modeling of the LoRaWAN

SF	T^{data} [s]	T^{ack} [s]	p_{equal}	$p_{EXPLoRa}$
7	0.051	0.041	0.166	0.487
8	0.102	0.072	0.166	0.243
9	0.185	0.144	0.166	0.135
10	0.329	0.247	0.166	0.076
11	0.659	0.495	0.166	0.038
12	1.318	0.991	0.166	0.019

Table II: Values of T^{data} , T^{ack} and SF distributions p . Payload of data packets is 10 bytes; ACKs have no payload.

standard considered in the simulator comes at the cost of a much larger computational time to assess the system performance. Indeed, for the same parameter set, the performance evaluation is basically instantaneous when employing the theoretical model, while each ns-3 simulation run takes in the order of tens of seconds, with execution times rapidly increasing when the traffic load, the number of devices and the number of required randomized runs grow.

The merit of the simulator is that it strives to be as realistic as possible, also taking into account some factors that are overlooked by the model for tractability reasons. For instance, the assumption of perfect orthogonality between transmissions employing different SFs is removed, and the simulator relies on the link-level model provided in [9] to determine the actual reception probability in case of overlapping transmissions, which also accounts for the capture effect.

The simulation setting is as follows.

- *Traffic load* – The number of EDs is fixed to 1200, and the EDs’ application layer is set to periodically generate packets to be transmitted by the MAC layer. The traffic load in the network is modified by varying the packet generation period. It is to be noted that this periodic traffic generation pattern is likely more realistic than the Poisson traffic assumed in the model. Nonetheless, the good match of simulation and analytical results confirms that the Poisson assumption is valid when the number of nodes is sufficiently large.
- *Channel allocation* – We consider the typical frequency allocation scheme for Europe, as reported in Tab. I. Therefore, the number of different frequency channels for UL is $C = 3$.
- *Duty cycle* – The simulator considers the DC limitations applied in the European region [12], which corresponds to setting $\delta_{SB1} = 99$ and $\delta_{SB2} = 9$ in the model.
- *Channel model* – Differently from the model, simulated LoRaWAN nodes experience a log-distance propagation path loss, as for an open-air scenario. Thus, farther devices will

suffer increased loss, and their performance will be penalized with respect to EDs that are close to the GW. Note that we do not include fast-fading components, which are supposed to be averaged out by the LoRa modulation, nor time-dependent variations in the channel, which remains constant throughout the entire simulation. Also, the channel is assumed to be symmetric, and DL transmissions will suffer the same impairments as in the UL.

- *SF distribution* – EDs are located around the single GW in a circular area of radius 2500 m, which allows for communications with any SFs with negligible channel error probability (in the absence of interference). Instead, the positions of the nodes are randomly picked at each simulation run. SFs are assigned uniformly (see Tab. II, p_{equal}). A different SF distribution (p_{EXPLoRa}) is considered in some scenarios, to evaluate the impact of this parameter on the different metrics.
- *Interference and capture effect* – To model interference, in the simulator we consider the collision matrix provided in [9] and the overlapping time between packets, as described in [23].⁴ A packet survives interference from a signal modulated with the same SF if its power is at least $CR_{dB} = 6$ dB higher than the colliding one. In order to provide a comparison with this scenario, in the analytical model we leverage the assumption of uniformly distributed EDs around the GW to compute the capture probabilities as in [16], which results in $\mathbb{W}^{GW} = 0.1796$, and $\mathbb{W}^{ED} = 0.5682$. We remark that different distributions of EDs around the GW can be modeled by adapting this derivation.

Since the GW implementation in the simulator attempts to emulate the behavior of a real device, a UL packet is successfully received when all the following conditions are satisfied:

- 1) The packet finds an available demodulator;
- 2) The packet's reception is not interrupted by DL transmissions;
- 3) Once the reception is finished, the packet was not corrupted by interference.

To count packets at the PHY layer coherently with the simulator implementation, the model's packet loss probabilities due to lack of demodulators (F_{NMD}), GW transmission (F_{GWTX}) and interference (F_{INT}) are plotted in the following section using, respectively, the following expressions:

- 1) $F_{NMD} = 1 - S^{demod}$;
- 2) $F_{GWTX} = E_i [S^{demod} \cdot (1 - S_i^{TX})]$;

⁴Note that, in the simulator, the capture event is determined also considering the partial overlapping of the colliding packets.

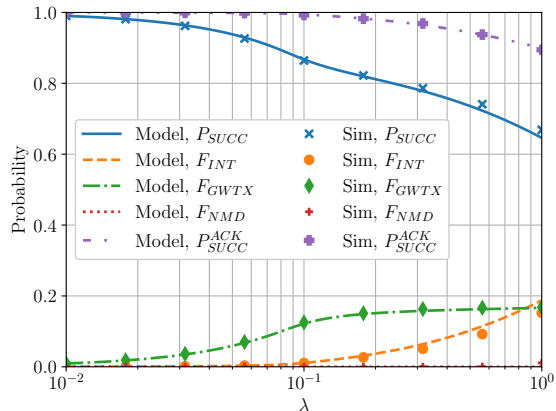


Figure 5: PHY-level performance with $m = 8$, $\alpha = 1$.

$$3) F_{INT} = E_i [S^{demod} \cdot S_i^{TX} \cdot (1 - S_i^{INT})];$$

by exploiting (6), (8), and (14), and where $E_i[\cdot]$ indicates the expectation over the distribution of SFs and S^{demod} the probability that, in the simulations, a packet can lock on an available demodulator.

VI. RESULTS

This section provides a comparison between the performance estimated with the proposed model and by the ns-3 simulator. Results are presented for both PHY and MAC layer, and the impact of the model's assumptions is shown to be mostly negligible, or at least acceptable. Finally, some results will showcase how the model can be used to gain insight on the behavior of the LoRaWAN technology in a quick and effortless way, analyzing the effects of various parameters on the performance of the network. In the plots of this section the analytical results are represented by lines, while markers correspond to simulation outcomes.

Fig. 5 shows the packet outcome probabilities at the PHY layer in a network employing confirmed traffic. Although obtained with different approaches, such probabilities are overall consistent, proving the effectiveness of the model.

The good match between model and simulation is also reflected in Fig. 6, which shows the CU and CD metrics for a network in which all EDs generate confirmed traffic ($\alpha = 1$), and for different values of m . Also in this case, the model results are quite close to those given by the simulations. Fig. 6a shows that the number of available transmissions helps the correct delivery of the message at the MAC layer, providing performance above 0.9 also for relatively high traffic levels, when an average of one packet per second is generated by the network at the application layer. The CD performance shown in Fig. 6b exhibits a similar behavior, but reaches much lower

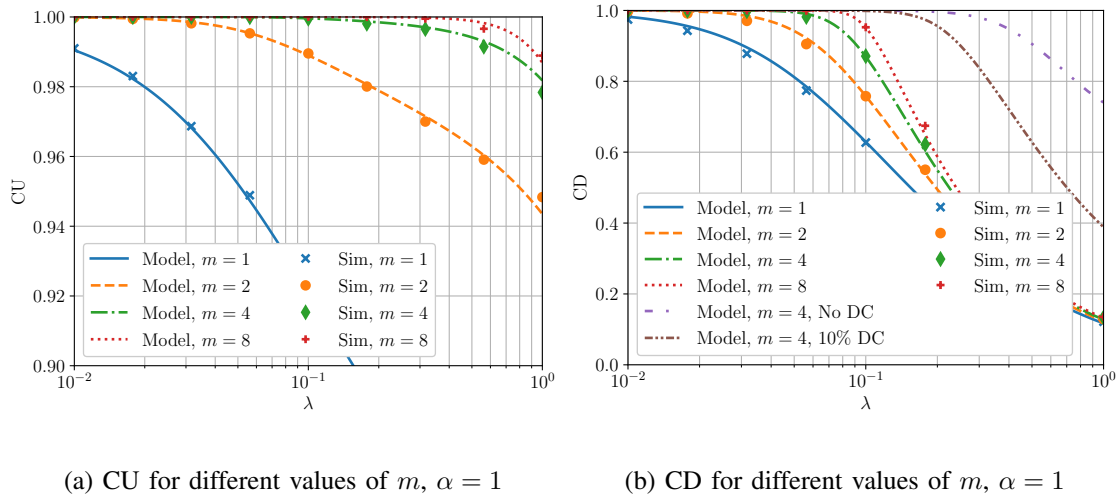


Figure 6: Comparison of model and simulation results in terms of CU and CD.

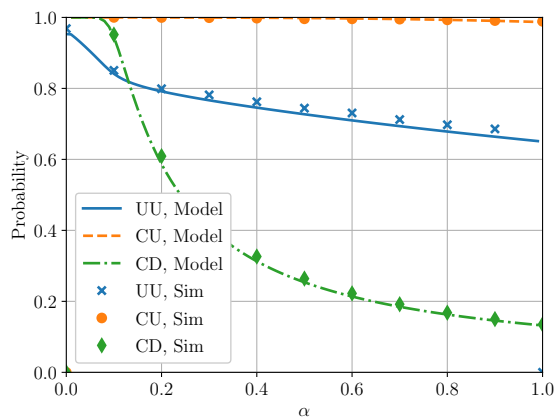


Figure 7: Performance when varying the fraction of confirmed traffic, with $\lambda = 1$, $m = 8$, $h = 1$.

values mostly because the rate of DL messages that the GW can generate is limited by the DC restrictions. The fact that this loss in performance is caused by the GW's DC is confirmed by the lilac dash-dotted line in Fig. 6b: to obtain these results, the DC restrictions were lifted by setting $\delta_{SB1} = \delta_{SB2} = 0$ in the model, producing markedly better results when compared to the corresponding green curve, where DC is enabled. Another example of the model's flexibility in considering also non-standard settings is given by the densely dash-dotted brown line, which represents the CD metric when $\delta_{SB1} = \delta_{SB2} = 9$, i.e., when transmissions in both sub-bands are subject to a DC of 10%. Although being an ideal setting, this case shows that even a small increase in the DC allowance in SB1 can yield considerable performance gains.

Fig. 7 compares simulation and theoretical results, in terms of UU, CU and CD, when different fractions of confirmed traffic are employed in the network. For this comparison, we set the network application layer packet arrival rate to $\lambda = 1$ pck/s, the maximum number of

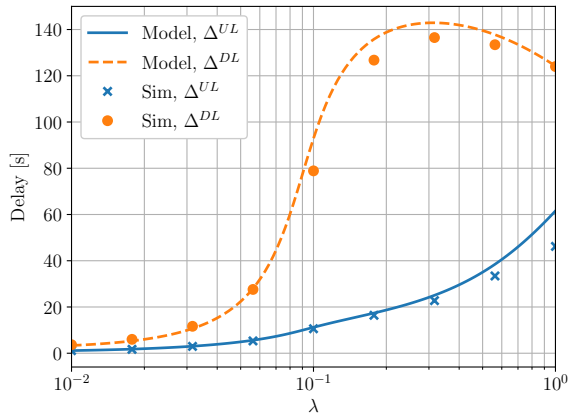


Figure 8: Delays for a confirmed traffic network, $m = 8$.

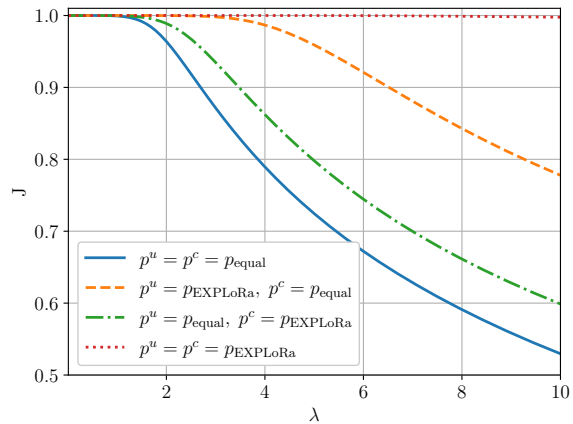


Figure 9: Fairness for different SF distributions when $m = 8, h = 8, \tau = 1, \alpha = 0.3$.

transmissions for confirmed traffic to $m = 8$, and the number of repetitions for unconfirmed traffic to $h = 1$. As the fraction of EDs employing confirmed traffic increases, the data delivery performance decreases for all the EDs, in particular for nodes employing unconfirmed traffic which do not have the chance of re-transmitting their packets. The match between the simulator and the model is confirmed to be excellent for all values of α .

The final metric that we evaluate through both model and simulation is the delay, as described in Sec. IV-G. Fig. 8 shows how delays generally increase with the traffic load, since more re-transmissions are needed to successfully deliver a packet. Note that for high values of λ the average ACK delay Δ^{DL} decreases: this is explained by the fact that devices employing higher SFs, (which may increase the average delay due to their longer inter-packet transmission times) heavily suffer from interference and are often dropped (unsuccessful packets are not considered in the delay computation). Although not shown here, it is worth noting that the model formulation makes it easy to extract per-SF metrics that can help troubleshoot the network configuration under study.

We now analyze how the fairness varies with the traffic load for different configurations of α , p^u and p^c . We consider the SF distributions p_{equal} and p_{EXPLoRa} as defined in Tab. III. The p_{EXPLoRa} distribution, first presented in [27], aims at equalizing the aggregate time on air of each group of devices employing the same SF to minimize the collision probability. In Fig. 9 we can observe that, when the SFs are uniformly allocated independently of the traffic type (i.e., $p^u = p^c = p_{\text{equal}}$), the fairness decreases for an increasing traffic intensity. Indeed, as the traffic grows, nodes employing lower SFs will suffer less from interference because of the

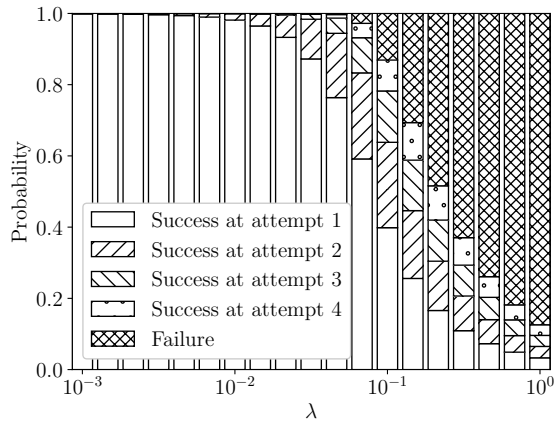


Figure 10: Distribution of re-transmissions, $m = 4$, $\alpha = 1$.

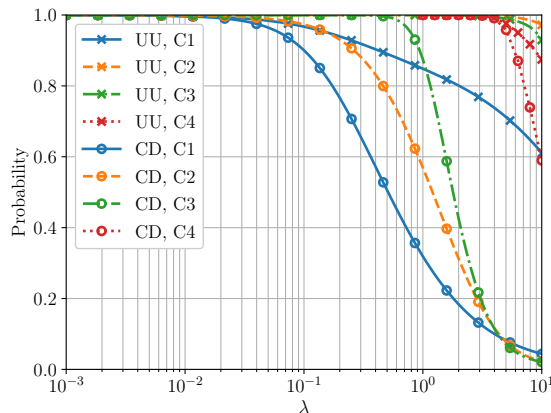
shorter transmission times. The fairness grows when $\alpha = 0.3$ and $p^c = p_{\text{EXPLoRa}}$, since with this configuration 30% of the generated packets will use lower SFs with higher probability, diminishing the channel and GW occupancy. However, since the traffic load is high and the fairness is measured on the uplink performance (UU and CU), the beneficial effect of allocating SFs according to the p_{EXPLoRa} distribution are more evident when it is used for most of the devices, i.e., the 70% of nodes employing unconfirmed traffic. Finally, the maximum fairness is achieved when the SFs are allocated using p_{EXPLoRa} both for p^u and p^c (dotted line in Fig. 9). Note that, when $\lambda \leq 1$, the load in the network is low enough to have $J = 1$ for every p^u, p^c , since the collision probability is low and the GW is not busy with ACK transmissions.

An example of insight that the analytical model can offer is presented in Fig. 10, which shows the fraction of traffic that achieves success after a certain number of re-transmission attempt for different traffic loads, derived from $P_{i,j}^{DL}$. This data, for instance, can be used to estimate the power consumption at the nodes: for low traffic loads the vast majority of MAC layer packet transmissions succeed with just one PHY layer transmission attempt. As the traffic load increases, the fraction of devices needing multiple re-transmissions to correctly receive an ACK correspondingly increases. After a certain point, packet reception fails with such a high rate that most EDs need to employ the maximum number of transmissions and, despite the high energy expenditure, still fail to receive an ACK from the GW.

Finally, we show how the model can be applied to investigate the impact of different network parameters on the performance. In the example of Fig. 11, 30% of the EDs employ confirmed traffic, and we show results obtained with the proposed mathematical model. The parameter configurations are summarized in Table III. Configuration C1 provides a baseline: priority is

Configuration	τ_1	τ_2	m	h	$p^u = p^c$
C1	1	1	1	1	p_{equal}
C2	0	1	1	4	p_{EXPLoRa}
C3	0	1	4	4	p_{EXPLoRa}

Table III: Configurations employed in Fig. 11

Figure 11: UU and CD performance for different network configurations, $\alpha = 0.3$.

given to DL transmission in both windows, devices employ a single transmission attempt for both confirmed and unconfirmed traffic, and SFs are uniformly distributed. In this case the curves have a shape similar to those shown in Fig. 6 for $m = 1$, but, since fewer devices require ACKs, the GW is able to receive more packets and profitably send replies, leading to better performance. To improve UU a second configuration (C2) considers the prioritization of ACK transmissions in RX2, where their reception suffers less interference. Moreover, unconfirmed packets are sent multiple times and we use $p^u = p^c = p_{\text{EXPLoRa}}$. This configuration provides a considerable improvement with respect to the UU metric, and some gains are also achieved in the CU performance. To improve also the results for confirmed traffic, a further step (configuration C3) is to set $m = 4$. This provides a significant improvement of CU, at the cost of a (minimal) decrease in UU performance. As a final step, we fully leverage the analytical model to identify the optimal parameter configuration (i.e., m , h , p_u and p_c) for each plotted traffic load, with the objective of maximizing the average of UU and CU. The red curves of this setting (C4) show how this optimization process enabled by the model can significantly improve the global performance of the network, significantly improving the CD performance at the price of a very small reduction in packet success rate for unconfirmed devices.

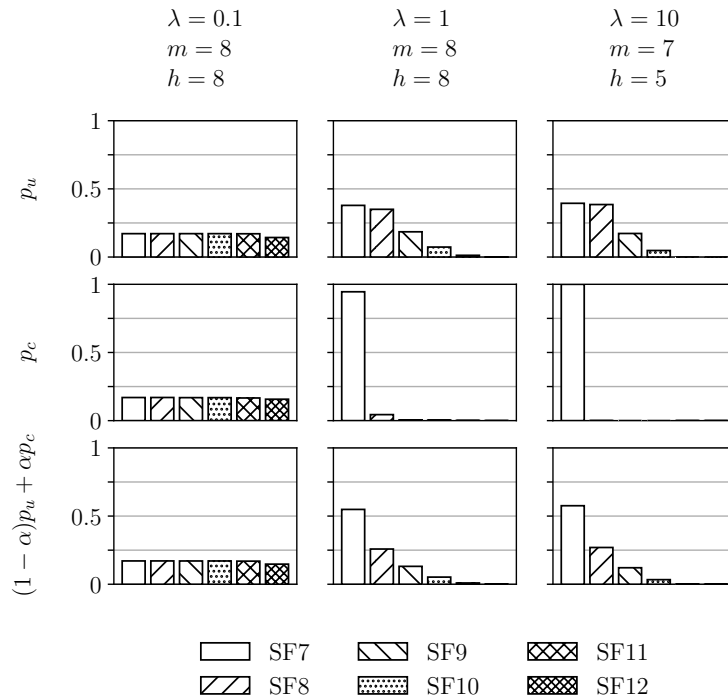


Figure 12: Optimal values of p_u , p_c , m and h as computed through model-driven optimization, for various values of λ .

The optimization problem that is solved to obtain configuration C4 is defined as:

$$\begin{aligned}
 & \max_{p_u, p_c} \quad \text{UU} + \text{CD} \\
 & \text{s.t.} \quad 0 \leq p_i^u \leq 1 \\
 & \quad \quad 0 \leq p_i^c \leq 1 \\
 & \quad \quad \sum_i p_i^u = 1 \\
 & \quad \quad \sum_i p_i^c = 1
 \end{aligned} \tag{38}$$

where we explore the entire space defined by m , h and λ , by solving (38) to find the best p_u and p_c , and finally pick the best solution for each λ . The search is performed using the trust region method as implemented by the `scipy` library, and we always set $p_i^u = p_i^c = 1/6$ as the initial parameter value for the algorithm.

Figure 12 displays the parameters of configuration C4 for some representative values of λ , showing p_u in the first row, p_c in the second row, and a combination of the two weighed on α on the third row. For a low value of generated traffic ($\lambda = 0.1$, first column), we see that the optimization stops almost immediately, yielding a distribution that is very similar to the initial

value of p_u and p_c . In this case, as can also be seen in Figure 11, since the traffic load is low the performance is indeed very good for high values of m and h , and needs little optimization of the SF distributions. For $\lambda = 1$, instead, the optimization process yields a more distinctive value of p_c , setting almost all devices to use SF7. This is motivated by the fact that, RX1 is set to employ the same SF used in the UL. Therefore, having most of the confirmed devices employ an SF as low as possible is advantageous, since it guarantees faster ACK transmissions in the DL and, as a consequence, shorter silent times imposed by the DC, and a larger set of devices can thus be served. Devices employing unconfirmed traffic, instead, are set to use a variety of SF values. Notably, the selected values are such that the aggregated distribution considering both unconfirmed and confirmed traffic (visible in the third row) takes a shape that is very similar to that of p_{EXPLORa} . This behavior is even more marked when $\lambda = 10$, with the notable difference that higher SF values are not used in the optimized network: this is because of the limited number of demodulators at the GW (a factor which is accounted for in our model). Indeed, although using all SF values would bring an additional gain, a packet with high SF value occupies a demodulator for quite a long time, increasing the probability that other incoming packets are dropped because of unavailability of reception chains at the GW. Finally, we note that m and h are consistently set to their maximum values (8 here) up to $\lambda = 1$. After this value, instead, it pays off to reduce the number of repetitions employed by both unconfirmed and confirmed EDs.

Although this analysis showcases the potential of the mathematical model to identify the optimal settings, an evaluation of the trade-offs associated to parameter configurations and their effect on other metrics of interest, such as delays and energy consumption, needs a deeper investigation, which we leave for future work.

VII. CONCLUSION

In this work, we presented a model for the performance evaluation of a LoRaWAN network in the presence of both confirmed and unconfirmed traffic, taking into account the influence of different settings of multiple network configuration parameters.

The model is able to capture both the PHY layer and MAC layer performance, and describes the multiple events that affect both UL packet reception and DL transmission: interference, capture effect, availability of demodulator, DC constraints, ongoing transmissions and receptions. We validated the model results with ns-3 simulations, showing the consistency among the two

sets of results. Finally, we presented some examples of how the model can be employed to analyze the effects of possible changes to the standard parameter settings, and to identify optimal configurations with minimum effort.

Several extensions of this work are possible. A first improvement to the model is the inclusion of multi-GW scenarios, where UL packets are potentially received by several GWs, and the network DL capacity is increased. A second aspect of interest is to leverage the proposed model to better investigate trade-offs among different network parameters in various scenarios, or when specific performance requirements are provided. A third possible improvement would involve characterizing the capture effect for non-uniform spatial distribution of the devices. Finally, a fourth direction is to employ the proposed model to identify optimal network settings when different metrics of interest are used as optimization functions, as we showed in the results section with some simple cases. We point out that the target of the model was to explore the capabilities of LoRaWAN networks, thus, in this work, we neglected some features of LoRa, such as the interference between overlapping packets modulated with different SFs. The model can be extended by including this, as well as other specific features of the LoRa technology. Such extensions are left for future work.

We remark that all figures contained in this paper, covering both model evaluations and simulation results, can be easily reproduced using the tool available at [25].

ACKNOWLEDGMENT

Part of this work was supported by MIUR (Italian Ministry for Education and Research) under the initiative "Departments of Excellence" (Law 232/2016).

REFERENCES

- [1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb. 2014.
- [2] Y. Yuehong, Y. Zeng, X. Chen, and Y. Fan, "The internet of things in healthcare: An overview," *Journal of Industrial Information Integration*, vol. 1, pp. 3–13, Mar. 2016.
- [3] N. Dlodlo and J. Kalezhi, "The internet of things in agriculture for sustainable rural development," in *International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, May 2015, pp. 13–18.
- [4] F. Chiariotti, M. Condoluci, T. Mahmoodi, and A. Zanella, "Symbiocity: Smart cities for smarter networks," *Transactions on Emerging Telecommunications Technologies*, vol. 29, no. 1, p. e3206, Jun 2018.
- [5] G. Bovenzi, D. Ciunzo, V. Persico, A. Pescapè, and P. S. Rossi, "IoT-enabled distributed detection of a nuclear radioactive source via generalized score tests," in *International symposium on signal processing and intelligent recognition systems*. Springer, Jan 2018, pp. 77–91.

- [6] F. H. C. dos Santos Filho, P. S. Dester, E. M. Stancanelli, P. Cardieri, P. H. Nardelli, D. Carrillo, and H. Alves, "Performance of LoRaWAN for Handling Telemetry and Alarm Messages in Industrial Applications," *Sensors*, vol. 20, no. 11, p. 3061, May 2020.
- [7] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: The rising stars in the iot and smart city scenarios," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 60–67, Nov 2016.
- [8] D. Croce, M. Gucciardo, S. Mangione, G. Santaromita, and I. Tinnirello, "Impact of LoRa Imperfect Orthogonality: Analysis of Link-Level Performance," *IEEE Communications Letters*, vol. 22, no. 4, pp. 796–799, Jan. 2018.
- [9] C. Goursaud and J. M. Gorce, "Dedicated Networks for IoT: PHY / MAC State of the Art and Challenges," *EAI Endorsed Transactions on Internet of Things*, vol. 1, no. 1, Oct. 2015.
- [10] Semtech Corporation, *SX1301 datasheet*, Jun 2014.
- [11] LoRa Alliance, "LoRaWAN™ 1.1 Specification," Oct. 2017.
- [12] —, "LoRaWAN™ 1.1 Regional Parameters," 2017.
- [13] 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, no. 9, pp. 34–40, Sep. 2017.
- [14] 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, Jan. 2017.
- [15] R. B. Sorensen, D. M. Kim, J. J. Nielsen, and P. Popovski, "Analysis of Latency and MAC-Layer Performance for Class a LoRaWAN," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 566–569, Oct. 2017.
- [16] D. Bankov, E. Khorov, and A. Lyakhov, "Mathematical model of LoRaWAN channel access with capture effect," in *IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct. 2017.
- [17] —, "LoRaWAN Modeling and MCS Allocation to Satisfy Heterogeneous QoS Requirements," *Sensors*, vol. 19, Sep 2019.
- [18] D. Croce, M. Gucciardo, S. Mangione, G. Santaromita, and I. Tinnirello, "LoRa technology demystified: From link behavior to cell-level performance," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 822–834, Feb 2019.
- [19] M. Heusse, T. Attia, C. Caillouet, F. Rousseau, and A. Duda, "Capacity of a LoRaWAN Cell," in *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2020, pp. 131–140.
- [20] F. H. Khan, R. Jurdak, and M. Portmann, "A Model for Reliable Uplink Transmissions in LoRaWAN," in *15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2019, pp. 147–156.
- [21] C. Delgado, J. M. Sanz, C. Blondia, and J. Famaey, "Batteryless LoRaWAN Communications Using Energy Harvesting: Modeling and Characterization," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2694–2711, Feb 2021.
- [22] M. Capuzzo, D. Magrin, and A. Zanella, "Mathematical Modeling of LoRa WAN Performance with Bi-directional Traffic," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018.
- [23] D. Magrin, M. Centenaro, and L. Vangelista, "Performance evaluation of LoRa networks in a smart city scenario," in *IEEE International Conference on Communications (ICC)*, May 2017.
- [24] A. Z. M. Z. Davide Magrin, Martina Capuzzo, "Proof of convergence of LoRaWAN model," *arXiv preprint arXiv:2108.02458*, 2021.
- [25] "lorawan-model," Available: <https://signetlabdei.github.io/lorawan-model/>.
- [26] D. Magrin, M. Capuzzo, and A. Zanella, "A Thorough Study of LoRaWAN Performance Under Different Parameter Settings," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 116–127, Jan 2020.
- [27] F. Cuomo, M. Campo, A. Caponi, G. Bianchi, G. Rossini, and P. Pisani, "EXPLoRa: Extending the performance of LoRa by suitable spreading factor allocations," in *IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2017, pp. 1–8.