Improving Downlink Scalability in LoRaWAN — **Source link**

Valentina Di Vincenzo, Martin Heusse, Bernard Tourancheau

Institutions: Polytechnic University of Turin

Topics: Scalability

Related papers:

- Understanding the Limits of LoRaWAN

- Low Power Wide Area Networks: An Overview

- Low Power Wide Area Network Analysis: Can LoRa Scale?

- Does Bidirectional Traffic Do More Harm Than Good in LoRaWAN Based LPWA Networks

- Reliable Group Communication in ATM Networks

Improving Downlink Scalability in LoRaWAN

Valentina Di Vincenzo, Martin Heusse, Bernard Tourancheau

Improving Downlink Scalability in LoRaWAN

Valentina DI VINCENZO
Politecnico di Torino
LIG CNRS, Grenoble Alps University
edu@first-last.com

Martin HEUSSE
LIG CNRS, G-INP
Grenoble Alps University
first.last@univ-grenoble-alpes.fr

Bernard TOURANCHEAU
LIG CNRS
Grenoble Alps University
first.last@univ-grenoble-alpes.fr

Abstract—**Although LoRaWAN offers many advantages, one of its main limitation is the insufficient downlink capacity in large-scale scenarios. This makes reliable communication impractical as, in LoRaWAN, reliability is achieved through the acknowledgement of confirmed messages. The contribution of this paper is to provide a comprehensive description of the problems associated with the presence of downlink traffic and to propose solutions to enable the use of acknowledgments for low to medium downlink load. The proposed mechanisms build on a more adequate and flexible choice of the gateway for each downlink transmission by the network server. Thus, they are entirely compatible with existing LoRaWAN technologies and deployments. Our simulation results show that the proposed solutions enable the successful realization of a higher percentage of LoRaWAN confirmed messages. Thus, the common practice of strongly discouraging the use of confirmed messages could be reconsidered to allow applications that require reliable communication.**

I. INTRODUCTION

Low Power Wide Area Network (LPWAN) solutions fill the gap between multi-hop sensor networks based on short range radio technologies (BLE or IEEE 802.15.4) [9] and cellular networks which are currently evolving to be able to serve massive amounts of clients [3] [14]. LoRaWAN is one of the most prominent LPWAN technology operating in the Industrial, Scientific and Medical (ISM) band, alongside with SigFOX and Weightless, among others. It allows long range communication between thousands of low power battery driven devices and always-on gateways in a cellular kind of architecture. In LoRaWAN, pure uplink traffic, from end-devices to gateways, is strongly favored. For this reason the existing literature focuses mostly on IoT best-effort applications, exhibiting scenarios with only unconfirmed messages, without taking into account downlink traffic even when it would have been necessary (e.g. [4] [13]). Indeed, sending sporadic data back to devices is an intrinsic necessity of LoRaWAN functioning, making downlink traffic a crucial and significant component for a proper scalability analysis of the technology. Besides, Maalel et al. [16] point out that mission-critical applications require a reliable data transport system. LoRaWAN implements a data retransmission scheme in order to improve transmission reliability by using acknowledgement messages (ACK) to guarantee data reception. Currently, network operators strongly restrict the generation of these kind of messages since the ACKs drain the network capacity, as shown in various studies (see Section VII). Thus, improving

TABLE I: LoRa transceiver (Semtech SX1276) parameters.

SF	7	8	9	10	11	12
Sensitivity [dBm]	-123	-126	-129	-132	-133	-136
DR [kbit/s]	3.4-5.5	2.0-3.1	1.1-1.8	0.6-1.0	0.3-0.5	0.2-0.3
min SNR [dB]	-7.5	-10	-12.5	-15	-17.5	-20

downlink communication efficiency and scalability seems to be an important need for LoRaWAN.

After an overview in Section II of the LoRa technology aspects that are relevant to our problematics, Section III identifies the challenges linked to the presence of downlink traffic. Section IV proposes three possible solutions to improve scalability issues. Section V addresses our simulation approach and the results are presented and discussed in Section VI. Before the conclusion, Section VII reviews the related works.

II. LoRa OVERVIEW

LoRa can refer to two different layers:

- **LoRa** is the physical layer, it defines the physical coding of data symbols. It exploits unlicensed ISM frequency bands and uses a Chirp Spreading Spectrum (CSS) radio modulation technique [23].
- **LoRaWAN** is the protocol stack (at the Media Access Control (MAC) and network layers) which exploits the LoRa physical layer. It is an open standard developed by the LoRa Alliance [22].

A. LoRa modulation

In LoRa, the frequency spreading is achieved by coding the information into the time offset applied to a base signal composed of a simple frequency ramp (or chirp) sweeping the entire bandwidth (BW). Indeed, the frequency bandwidth of the chirp is equivalent to the spectral bandwidth of the signal. Each chirp spans 2^{SF} samples at the *BW* sampling rate and there are 2^{SF} different time offsets, which allows to convey SF bits. As a result, LoRa trades data rate (DR) for sensitivity within a fixed channel BW, as shown in Table I for the 125 kHz bandwidth modulations. Notice that an error correcting code (generally of coding rate (CR) 4/5) introduces some redundancy at the bit encoding level. The chirps at different DR are orthogonal with each other. This allows the receiver to simultaneously receive multiple transmissions on different SFs as described in the specifications [21].

B. LoRaWAN

LoRaWAN presents a cellular star-of-stars topology [26] with three main entities:

- **The end-device** emits LoRa packets towards a server each time a specific event triggers the transmission.
- **Gateways** receive LoRa packets from all LoRa end-devices sending traffic in their reception range. Each gateway is registered to a network-server to which it transparently forwards all the received traffic through an IP connection.
- **Network-Server** is in charge of all the decisions and parameters adaptation. Several gateways can be registered to the same server and they can receive the same packet more than once. After a de-duplication operation, the packet is forwarded to the application.

In order to allow end-devices to receive downlink frames, three different classes are defined based on energy consumption:

- **Class A - All end-devices** are booted in this mode. It allows the lowest energy consumption possible since the end-device only gets to receive downlink traffic in one of two reception slots after each transmission. Every time an uplink frame is sent, a first receive window, RX1, is opened by the end-device RECEIVE_DELAY1 seconds after the end of the transmission (default value 1 second) with the same SF used for the uplink frame increased by a fixed amount (default 0) and using the same carrier frequency. If a preamble is detected, the radio receiver stays active for the downlink frame reception. If no downlink frame is received in RX1, the end-device opens a second receive window, RX2, after RECEIVE_DELAY2 seconds (default value RE-CEIVE_DELAY1 + 1 seconds) with preset parameters (default values 869.5MHz/SF12/125kHz).

Fig. 1: Class A end-devices receive slot timing.

- **Class B - Beacons.** A Class A end-device can decide to switch to Class B. The purpose is to make an end-device available to receive downlink frames at other predictable times, in addition to RX1 and RX2.
- **Class C - Continuous listening** allows devices to constantly be in receive mode, except when an uplink frame has to be sent. Obviously, this mode of operation is not suitable for devices with energy constraints.

The most relevant LoRa message fields for this work relate to the generation and management of downlink traffic:

- **Adaptive Data Rate**: LoRaWAN allows end-devices to adapt the DR in order to increase the battery life and maximize the network capacity. When the ADR bit is set,

the network can control the DR through the associated MAC commands *LinkADRReq* and *LinkADRAns*. If an end-device uses a DR which is greater than its default one, it needs to periodically validate that its uplink frames are still received by the network server. Thus, if the device has not received a downlink frame after ADR_ACK_LIMIT (typically 63) uplink frames, the ADRACKReq bit is set. This informs the network server to send a downlink frame as soon as possible. If after ADR_ACK_DELAY (e.g. 32) more transmissions, the end-device has still not received a downlink, it switches to the next lower DR to try regaining connectivity.

- **ACK**. A CONF is a frame with a bit set in the header that requires an acknowledgement from the receiver. If a CONF is sent by the end-device, the network server must send a downlink frame with the ACK bit set in one of the end-device RX windows. If an acknowledgement frame is not received after a number of retransmissions (default value 0), the end-device may try to regain connectivity by lowering its DR to send the message again, until eventually discarding it. It is worth noticing that acknowledgments are sent by the server only in response to the latest CONF received and are never retransmitted.

Moreover, LoRaWAN offers several MAC commands to the network server and the end-devices, in order to adapt various physical layer parameters and to assess the status of devices or network links. Each time such MAC command is sent by the end-device, it must be answered/acknowledged by the server and thus, a downlink frame is involved.

Also, in order to participate in a LoRaWAN network, each end-device has to be activated in one of the two following manners:

- **Over-The-Air-Activation** (OTAA) requires an end-device *join-request* uplink frame and a network server *join-accept* downlink frame if the end-device is permitted to join the network. This type of activation is the most secure.
- **Activation By Personalization** (ABP) directly ties an end-device to a specific network, bypassing the join procedure. The end-device is assumed to have the required information for participating in the network when started.

III. DOWNLINK TRAFFIC: PROBLEMATICS

In a nutshell, downlink frames arise for a variety of reasons:

(a) Reception of CONF and subsequent ACKs;
(b) Management functions required by LoRaWAN, e.g. the OTAA join procedure for devices;
(c) MAC commands for parameters adaptation and link connectivity testing, e.g. LinkADRReq, LinkCheckAns, etc.;
(d) In response to frames with ADRACKReq bit set;
(e) Events notification from IoT applications (e.g. software update, sampling interval change, ...).

Even if CONFs are currently not recommended due to their negative impact on the network, there will be downlink traffic for reasons (b), (c), (d), and possibly (e). Hence, it is important

to understand which are the problems related to downlink frames along with the normal uplink traffic and how this affects network capacity.

Below, we analyze three limiting factors for downlink frames transmission.

A. Duty-Cycle Saturation

LoRaWAN operates in the ISM band which is regulated by the ETSI standard [11]. In order to be compliant, the LoRaWAN specification exclusively uses duty-cycled transmissions with an ALOHA access method [15]. In Europe, LoRaWAN operates in the 863-870 MHz band, for which the standard defines five sub-bands with per sub-band limitations. In those bands, the duty cycle ratios ($DutyCycle_{subband}$) range from 0.1% to 10%. Each time a frame is transmitted, its emission timestamp and its Time On Air (ToA) are recorded by the emitter for each transmission sub-band. The ToA defines the transmission time of a frame and it depends on the frame size, SF, BW and CR [20]. Then the same sub-band cannot be used again for the subsequent time off ($TOff_{subband}$) seconds, where

$$TOff_{subband} = \frac{ToA}{DutyCycle_{subband}} - ToA \qquad (1)$$

Both end-devices and gateways comply to this limitation. Therefore it is evident that, in scenarios with downlink traffic, gateways can potentially become the bottleneck of the network by saturating their $DutyCycle_{subband}$ and hence exhausting their ability to forward downlink frames to the numerous end-devices. Indeed, taking into consideration class A end-devices (the most used), the gateway must send the downlink frame exactly at the beginning of one of the two RX windows, in order to allow the radio receiver of the end-device to detect the downlink preamble. If the duty-cycle is saturated for the sub-band corresponding to the downlink transmission channel, the gateway will not be able to forward the downlink frame. Next, the end-device will perform the retransmission of the CONF and possibly end up switching to a higher SF, with a greater and greater radio channel footprint.

B. The gateway radios are Half-Duplex

Due to the lack of frequency band separation between uplink and downlink communication, LoRa gateways operate in half-duplex mode, i.e. they cannot receive and transmit at the same time. When a frame is sent, all ongoing receptions are aborted and no uplink frame can be received for all the duration of the downlink frame ToA, called the *no-reception window*. The presence of downlink traffic in a large network will consequently cause an overall decrease of the end-devices PDR. We thus identify two different modes related to the gateway's ability to receive frames:

- "*free*" when the channel is used for receiving uplink frames. In this mode the gateway is available for reception and all received uplink frames are forwarded to the network server.
- "*busy*" when the transmission direction is switched and the gateway sends a downlink frame. When in this mode, the gateway is unable to receive any uplink frame.

C. Downlink frames sequential sending

The modulation profile (SF, CR, BW, ..) and the emission timestamp of a downlink frame are computed by the network server at the downlink frame scheduling phase. A LoRa gateway can only have one frame programmed for departure at a time, when the internal clock reaches the downlink timestamp value. Thus, downlink frames are buffered in a Just In Time queue and sent exclusively sequentially. This hardware constraint exacerbates the problem caused by the gateway's half-duplex mode and speeds up the duty-cycle saturation process for downlink traffic.

IV. PROPOSED SOLUTIONS

We propose various strategies to tackle the identified downlink traffic issues.

A. Multi-Gateway

One way of obviating the half-duplex nature of the gateways is to change the network architecture. Indeed, by deploying additional gateways (even with small coverage zones), many uplink frames which overlap in time with downlink transmissions could still be received by free-gateways. In this scenario, the no-reception window of a busy-gateway is covered by all the other free-gateways able to receive uplink frames and forward them to the network server. Note that the base chirps used for uplink and downlink transmissions are complex conjugate of each other. They are orthogonal, so that downlink transmissions from distant gateways have little impact on uplink reception.

B. Parallel sending and downlink frames combination

The current hardware design of LoRa gateways allows the transmission of a single frame at a time. Nevertheless, it is possible to imagine a hardware design able to overcome this restraint, since it is not inherent the technology. The idea here is to not only allow frames emissions in parallel on different channels, but also simultaneously on the same channel with different orthogonal spreading factors, similarly to what is done in UMTS cellular networks. Indeed, if LoRaWAN gateways had the ability to send more downlink frames simultaneously, the no-reception window and the $TOff_{subband}$ would improve considerably. Two factors have to be taken into account in the implementation of the downlink frames parallel sending:

- Downlink frames combination requires that the receive windows of both candidate receivers overlap. This can be made more likely by lengthening the RX duration, as explored by Centerano and Vangelista [7].
- The gateway must consider also the restriction on the maximum Effective Radiated Power (ERP) imposed by the ETSI regulation. All the sub-bands used by LoRa are limited to a maximum radiated power of 14 dBm, except for the default sub-band of RX2, characterized by an ERP limitation of 27 dBm. As a consequence, to stay within the ERP limitation, it is necessary to compute a correct transmission power profile when combining several downlink frames on the same sub-band.

C. Balanced Gateway Selection Algorithm

In a multi-gateway scenario, the network server could (and actually should) possibly receive the same uplink frame more than once, forwarded by different gateways. When the network server needs to send a downlink frame to an end-device, it has to select the gateway that will perform the transmission. The LoRaWAN specifications do not provide any information or recommendation on how the network server should perform such operation. One possible selection algorithm, adopted by an open-source LoRaWAN network-server, the LoRaServer project [5], is based on the Signal to Noise Ratio (SNR). This is a good way to estimate the wireless link quality [24] assuming the same bidirectional transmission conditions. In this SNR-based algorithm, the selection of the gateway only depends on the information carried by the duplicates received from the targeted end-device. In the de-duplication phase, the server records the gateway that forwarded the frame with the best SNR value.

Although the radio link quality matters, it is clear that if the gateway is in *Toff* for RX1 and RX2, then it is for sure unable to send the downlink. As emphasized in Section III, the duty-cycle of a gateway which serves a large number of end-devices may quickly reach saturation and the gateway would then miss the RX windows. Moreover, if the selected gateway is consistently receiving a large amount of uplink traffic, the number of missed uplink frames during the downlink transmissions would most probably also be significant. Thus, with such a naive selection algorithm, the gateway load is not taken into consideration and it mechanically introduces notable frame loss. In order to better balance and spread the downlink traffic throughout the network, the gateway selection algorithm should also take into account the effective ability of the gateway to forward the downlink, by checking its duty-cycle saturation and its already scheduled downlink frames, in order to avoid overlaps, similarly to what is currently done by The Things Network [25]. This multicriteria gateway choice algorithm attempts to strike a balance between trying to use the best radio link and avoiding losses due to duty-cycle limitations and to the downlink frames sequential sending.

V. SIMULATION DATA & EVALUATION APPROACH

In order to validate the proposed solutions experimentally, it would be necessary to deploy not just some gateways and a network server, but also a considerable number of end-devices. Indeed, the network server can only interact (e.g. sending downlink frames) with registered end-devices, whereas all the traffic received from non-registered devices is ignored. We chose the pragmatic approach to implement a simulator to evaluate and quantify the improvements brought by the proposed solutions. But instead of having a traffic generation model which follows a Poisson distribution as in many studies [2], [12], our simulations exploit real traffic traces to observe the behavior of the network when it is virtually exposed to various percentages of downlink frames, relatively to the uplink traffic. In order to collect real traffic traces, four Semtech LoRa PicoCell Gateways [19] were set up

and attached to Raspberry Pi boards. The WalT platform [6] greatly facilitated the deployment. As shown in Table II, the gateways positions were meant to produce a configuration where a gateway serves a large number of nodes (GW1) and at least part of the received traffic is shared with other gateways, in order to have some duplicates.

A. Virtual downlink traffic generation

In real traffic, usually most, if not all, uplink frames are unconfirmed frames. In order to analyze the network reaction to various percentages of downlink frames, we vary the proportion of CONF virtually. In practice, the simulator randomly selects in the considered trace an unconfirmed frame (and all its duplicates) and marks it (them) confirmed. This is repeated until a new trace is produced with the targeted percentage of CONFs.

B. Downlink frames scheduling algorithm

In the simulation, a downlink frame can be triggered by the reception of either an unconfirmed message with the ADRACKReq bit set or a CONF. In order to send a downlink, the simulator has to select the gateway that will take care of the downlink transmission: gateways are ordered by the SNR value of the forwarded uplink frame and the "best" gateway is initially selected.

Two conditions are necessary so the downlink frame can actually be scheduled on the selected gateway in any of the two RX windows.

1) **Sequential sending**: the downlink frame must not overlap for all its duration with the ToA of any other downlink frames already scheduled on this gateway.
2) **Duty-cycle**: the $TOff_{subband}$ must not overlap with the $TOff_{subband}$ of any message already scheduled in that sub-band for the gateway.

A first attempt is performed to schedule the downlink frame in RX1, as shown in Figure 2. If one of the two criteria is not met, a second attempt is performed in RX2. If again the downlink frame cannot be scheduled in RX2, then we distinguish two algorithms:

(a) **SNR-based algorithm:** the downlink frame (and thus the confirmed message associated) is considered lost and we record which unsatisfied criteria caused the loss.

(b) **Balanced algorithm:** other attempts in RX1 and RX2 are performed with the second best gateway in terms of SNR, and so on. Only when the downlink frame cannot be scheduled at any of the gateways that forwarded the corresponding uplink frame, it is considered lost.

TABLE II: Gateways set up for real traces collection.

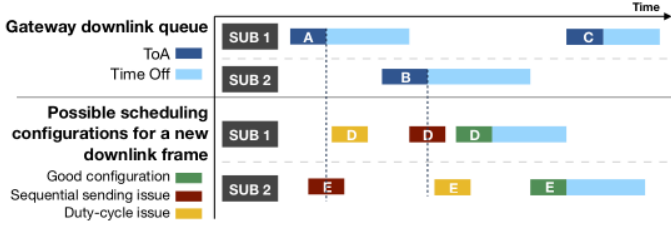| GW	Displacement	Received traffic
1	Outdoor	100%
2	Outdoor	52%
3	Indoor	6,5%
4	Indoor	4,7%

Fig. 2: Scheduling illustration for a new downlink frame (D or E). Downlink frames A, B and C are already scheduled.

VI. SIMULATION RESULTS

A. Single-gateway architecture

One of the contributions of this paper is to give clear insights about the problems related to downlink traffic.

In this experiment, the simulator is fed with the real traffic trace of the main gateway (GW1). The percentage of CONF varies virtually from 0% to 100% in order to understand how the network reacts. The captured traffic is intense enough to saturate the downlink channels, as long as enough frames are virtually changed to confirmed.

Each frame loss shown is due to one of the causes exposed in Section III. Thus, these frames are surely lost and would have been received if downlink traffic was not introduced in the network.

In Figure 3, CONF can be considered lost due to:

- **ACK lost**: If one of the two criteria detailed in Section V is not satisfied for both RX1 and RX2, the ACK is lost. The unmet criterion is registered only for the failure in RX2. Thus, the loss can be caused by:
  - Failure in RX1 and duty-cycle saturation of RX2 subband (in red)
  - Failure in RX1 and downlink frames sequential sending for RX2 start time (in yellow)
- **Half-duplex mode**: the CONF was supposed to be received while the gateway is sending, for this reason it is lost (in cyan).

Unconfirmed messages can be lost only because the gateway radio is half-duplex – and busy transmitting (in purple).

The sum of all the four loss components gives the total frame loss over the percentage of confirmed messages. Figure 3a highlights three main factors:

1) With a low percentage of CONF, most losses are related to unconfirmed messages. Indeed, if the downlink load is low, it does not congest the gateway and ACKs can be sent. On the other hand, every time a downlink frame is sent, all the uplink frames received during that period are lost.

2) From 20% of CONF and on, the main cause of frame loss is duty-cycle saturation. If all the frames are confirmed (100% of CONF), 8% of them are not received by the gateway because of its half-duplex nature, and 76% of the requested ACKs cannot be transmitted. In this case, the frame loss reaches 86%: almost all frames are lost.
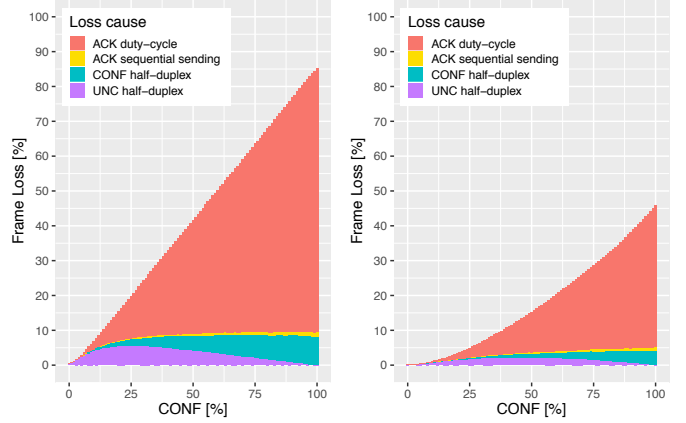
(a) Single-gateway implementation.

(b) Quad-gateway implementation.

Fig. 3: Frame loss due to downlink traffic.

3) The sequential sending of downlink frames causes no more than 2% of losses even with 100% CONF. This can be explained by observing that a heavier traffic load than the one deployed here would correspond to a higher downlink frames demand. With more downlink frames requested, the probability of overlapping with other downlink frames already scheduled in the gateway would surely increase. Since the main ACK loss is due to the duty-cycle saturation, the parallel sending solution will not produce a tangible improvement with these data (its validation is left for future work).

B. Multi-gateway architecture

In order to decrease the frame loss due to the gateway half-duplex mode, a second simulation was performed with a multi-gateway implementation. The simulator was fed with a real traffic trace file that combines frames received by the main gateway and duplicates received in the same time window by three other gateways. Figure 3b shows the improvement brought by a quad-gateway implementation with the SNR-based selection algorithm. The frame loss is almost halved. Indeed, all the frames that are lost while deploying just one gateway, are received by the other free-gateways. Moreover – whereas in a single-gateway implementation all the downlink frames are sent through the same gateway– in a multi-gateway scenario, downlink traffic is spread among several gateways, decreasing the congestion. As a result, more downlink frames can be sent in RX1 and the RX2 dedicated sub-band is less (or more gradually) saturated.

C. Gateway selection algorithms

Here, despite the use of more gateways, frame loss for higher proportion of confirmed messages is still important. This is a direct consequence of the behavior of the gateway selection algorithm which does not attempt to spread downlink traffic in any way by always and only selecting the gateway with best reception conditions.
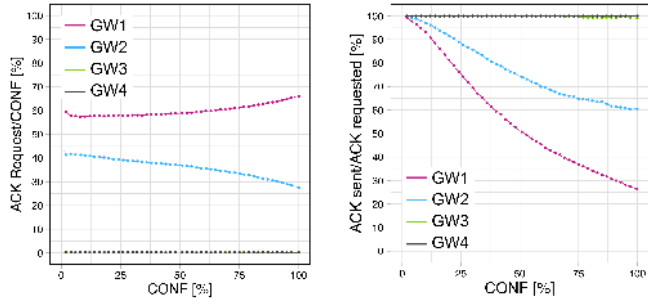
(a) Percentage of ACK requested per gateway under total number of CONF frames.

(b) Percentage of ACK sent on ACK requested per gateway.

Fig. 4: Gateways behavior under SNR-based selection algorithm.

Fig. 5: Frame loss percentage under different algorithms and number of deployed gateways.

This concentration of downlink traffic is clear in our simulations: Figure 4a shows that GW1 is selected at least 60% of the time when an ACK has to be scheduled; in Figure 4b it is possible to see that, as a consequence of such high downlink frame request load, the effective percentage of ACKs sent by the gateway significantly decreases with the increasing load. With 100% of CONF, GW1 is only able to send 24% of the requested ACKs. On the other side, two gateways (GW3 and GW4) are almost never used, while GW2 is complementary to GW1. To better balance the load of the downlink traffic, the balanced gateway selection algorithm is tested on the same trace. We show the improvements introduced by the balanced algorithm compared to the SNR-based one with their respective frames loss. Figure 5 shows that, in a quad-gateway implementation (in green), the balanced algorithm decreases the frame loss by 25% compared to the SNR-based algorithm and by 66% compared to the single-gateway architecture (in red), producing losses that never exceed 20% of the total traffic. The main reason of this improvement is the better distribution of the downlink traffic. ACK loss, which was the main cause of frame loss for high percentage of CONF, is strongly decreased. Unlike before, a downlink frame is not scheduled on the gateway with the best SNR if it is duty-cycle saturated for the given sub-bands or it overlaps with other already scheduled downlink frames. The algorithm tries the second best gateway in terms of SNR, and so on. Obviously, ACKs can still be lost due to other factors, especially if transmitted by a gateway that presents a lower SNR value. But in fact, the downlink channel is less challenging than the uplink as in all cellular networks: nodes are in view of less contending entities than the gateways, which use a specific modulation and are generally placed so as to receive from as many nodes as possible.

VII. RELATED WORK

Downlink traffic is arousing interest only recently. Most of the studies on the topic mostly focus on showing its negative impact on the network, usually identifying the duty-cycle as the main or single problem.
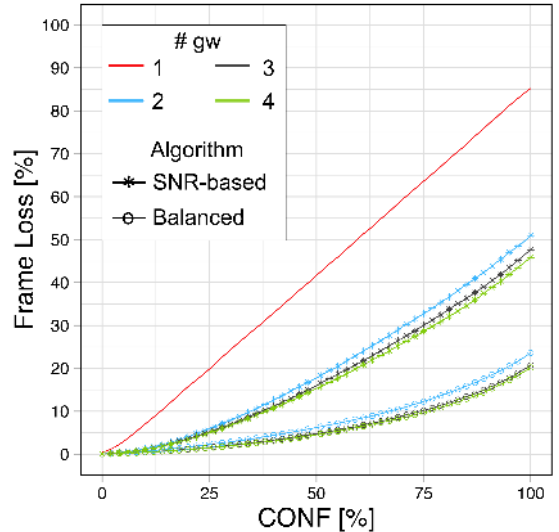
Adelantado et al. [1] conduct an analysis of LoRaWAN in order to identify the main problems of the technology. Among other issues, the duty-cycle saturation is recognized as one of the limits of LoRaWAN, especially for the development of ultra-reliable applications which, as such, require acknowledgments.

Pop et al. [18] address the bidirectional traffic problem by introducing LoRaWANSim simulator, an extension of LoRaSim that includes downlink frames and retransmissions. Some of the findings are relative to the aggressive data-rate back-off approach during retransmission recommended in the old version of LoRaWAN specification, and hence they are not relevant anymore. On the other hand, they correctly identify the scalability problems that arise when a large number of nodes request ACKs and the duty-cycle limitation issue.

An empirical evaluation of the downlink traffic impact is conducted by Mikhaylov et al. [17]. The results show the strong negative effect on the performance of uplink frames, consistent with the previous studies. Nevertheless, the explanation behind this phenomenon does not take into account the half-duplex nature of the gateway and the I/Q inversion performed by gateways when transmitting uplink and downlink frames, in order to avoid their collision.

The aim of Centenaro et al. [8] is to investigate the impact of downlink frame feedback in a multi-gateway architecture using an event-driven simulator. In order to have some scheduling flexibility on the network side, one of the modifications of the model is to have receive windows one second longer than the actual ones. This significant change imposes that each RX window stays open for an unreasonable amount of time and the resulting downlink frame scheduling process is completely different from the real one. However, the obtained results show again the negative impact of downlink frames without providing additional reasons besides the duty-cycle issue.

One of the main contributions to the downlink frame issue is given by Van et al. [10] by performing a downlink traffic analysis using the ns-3 network simulator. The authors clearly explain the reasons of gateways congestion by highlighting the duty-cycle limitation, whereas, the half-duplex problem is merely mentioned. A multi-gateway architecture is proposed and improvements are evaluated in terms of lower duty-cycle saturation and better distribution of the workload through the gateways. Yet, neither the explanation, nor the proposed solution, take into consideration the gateway selection algorithm. In their implementation, the server simply tries to schedule an ACK in the first available gateway among the pertinent gateway(s).

VIII. Conclusion

This paper presents a comprehensive analysis of the impact of downlink frames on the network capacity. The roots of the problem are identified and checked on a real trace file: the fact that gateway transmissions are half-duplex (either sending or receiving) causes up to 8% of frame loss; the downlink frame sequential sending at the gateways - by which only one downlink frame can be scheduled at a given time - produces up to 2% of frame loss due to downlink frame timing overlapping with other already scheduled ones; and the duty-cycle limitation that induces up to 76% of frame loss due to the saturation of the downlink frame sub-band. We propose three solutions in order to mitigate such effects: a multi-gateway implementation, downlink frames parallel sending and a balanced gateway selection algorithm. Two of them are tested by the implementation of an event-driven simulator fed with captured real traces. From an initial 86% of frame loss with a single-gateway architecture, the proposed solutions provide an enhancement of 66% in a quad-gateway scenario in presence of a high downlink communication load corresponding to 100% confirmed messages. Moreover, for low and medium downlink traffic loads, frame loss never exceeded 5% thanks to the balanced gateway selection algorithm, enabling the deployment of applications that require such confirmed data communication.

References

[1] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, and T. Watteyne. Understanding the limits of LoRaWAN. *IEEE Communications Magazine*, 55(9):34–40, Sept 2017.

[2] D. Bankov, E. Khorov, and A. Lyakhov. Mathematical model of LoRaWAN channel access. In *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks*, pages 1–3, June 2017.

[3] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski. Five disruptive technology directions for 5G. *IEEE Communications Magazine*, 52(2):74–80, February 2014.

[4] M. C. Bor, U. Roedig, T. Voigt, and J. M. Alonso. Do LoRa low-power wide-area networks scale? In *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 59–67, New York, NY, USA, 2016. ACM.

[5] O. Brocaar. LoRa server, open-source LoRaWAN network-server. www.loraserver.io.

[6] P. Brunisholz, E. Duble, F. Rousseau, and A. Duda. WalT: A reproducible testbed for reproducible network experiments. In *2016 IEEE INFOCOM Workshop CNERT*, pages 146–151, April 2016.

[7] M. Centenaro and L. Vangelista. Boosting network capacity in LoRaWAN through time-power multiplexing. In *Proc. IEEE PIMRC*, pages 1–6, Bologna, Italy, Sep. 2018.

[8] M. Centenaro, L. Vangelista, and R. Kohno. On the impact of downlink feedback on LoRa performance. In *IEEE Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–6, Oct 2017.

[9] 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*, 23(5):60–67, Oct 2016.

[10] F. V. den Abeele, J. Haxhibeqiri, I. Moerman, and J. Hoebeke. Scalability analysis of large-scale LoRaWAN networks in ns-3. *IEEE Internet of Things Journal*, 4(6):2186–2198, Dec 2017.

[11] European Telecommunications Standards Institute. Electromagnetic compatibility and radio spectrum matters (ERM); short range devices (SRD); radio equipment in the 25 MHz to 1000 MHz with power levels up to 500 mW. Tech. Rep. EN300 4.1, European Telecommunications Standards Institute, 2013.

[12] O. Georgiou and U. Raza. Low power wide area network analysis: Can LoRa scale? *IEEE Wireless Communications Letters*, 6(2):162–165, April 2017.

[13] V. Gupta, S. K. Devar, N. H. Kumar, and K. P. Bagadi. Modelling of IoT traffic and its impact on LoRaWAN. In *IEEE Global Communications*, pages 1–6, Dec 2017.

[14] S. Lien, K. Chen, and Y. Lin. Toward ubiquitous massive accesses in 3GPP machine-to-machine communications. *IEEE Communications Magazine*, 49(4):66–74, April 2011.

[15] LoRa Alliance Technical Committee Regional Parameters Workgroup. LoRaWAN^{TM} 1.1 specification. Tech. rep., LoRa Alliance, January 2018.

[16] N. Maalel, E. Natalizio, A. Bouabdallah, P. Roux, and M. Kellil. Reliability for emergency applications in internet of things. In *IEEE International Conference on Distributed Computing in Sensor Systems*, pages 361–366, May 2013.

[17] K. Mikhaylov, J. Petjjrvi, and A. Pouttu. Effect of downlink traffic on performance of lorawan lpwa networks: Empirical study. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2018.

[18] A. Pop, U. Raza, P. Kulkarni, and M. Sooriyabandara. Does bidirectional traffic do more harm than good in LoRaWAN based LPWA networks? In *2017 IEEE Global Communications Conference, GLOBECOM 2017, Singapore, December 4-8, 2017*, pages 1–6, 2017.

[19] Semtech Corporation. LoRa PicoCell gateway user guide. www.semtech.com.

[20] Semtech Corporation. Sx1272/3/6/7/8 LoRa modem design guide, 2013.

[21] Semtech Corporation. LoRa modulation basics, May 2015. www.semtech.com.

[22] N. Sornin and A. Yegin. LoRaWAN^{TM} 1.1 specification. Tech. rep., LoRa Alliance, October 2017.

[23] A. Springer, W. Gugler, M. Huemer, L. Reindl, C. C. W. Ruppel, and R. Weigel. Spread spectrum communications using chirp signals. In *IEEE/AFCEA EUROCOMM 2000. Information Systems for Enhanced Public Safety and Security*, pages 166–170, May 2000.

[24] W. L. Tan, P. Hu, and M. Portmann. SNR-based link quality estimation. In *IEEE Vehicular Technology Conference*, pages 1–5, May 2012.

[25] The Things Network. Downlink configuration router, 2018. www.thethingsnetwork.org/docs/network.

[26] L. Vangelista, A. Zanella, and M. Zorzi. Long-range IoT technologies: The dawn of LoRa^{TM}. In *Future Access Enablers for Ubiquitous and Intelligent Infrastructures*, pages 51–58. Springer, 2015.