Adaptive mitigation of the Air-Time pressure in LoRa multi-gateway architectures

Francesca Cuomo*,†, Manuel Campo‡, Enrico Bassetti*, Lorenzo Cartella*, Federica Sole*, Giuseppe Bianchi†‡

*University of Rome “La Sapienza”, Italy - †University of Rome “Tor Vergata”, Italy - ‡CNIT, Italy

Abstract—LoRa is a promising technology in the current Internet of Things market, which operates in un-licensed bands achieving long-range communications and with ultra power devices. In this work we capitalize on the idea introduced in [1], i.e. balance the Air-Time of the different modulation spreading factors (SF), and adapt it to operate in a typical metropolitan scenario comprising multiple gateways (GWs) interconnected to a same network server. Our proposed approach, named ADaptive Mitigation of the Air-time pressure in IORa (AD MAIORA), relies on a suitable measure of the per-spreading-factor load at each GW - quantified by means of a so-called pressure table -, and on a relevant heuristic algorithm which attempts to balance such a per-SF-pressure. Especially in cases of very loaded scenarios, where a high number of nodes insist on the same GWs, the use of AD MAIORA shows significant performance gains, up to a factor of 5 improvements with respect to the legacy LoRaWAN’s Adaptive Data Rate.

Keywords—Low power wide area networks; Internet of Things; LoRaWAN; Spreading Factors; Multi-Gateway.

I. INTRODUCTION

LoRa is a promising technology in the current Internet of Things (IoT) market. It operates in the ISM band using a proprietary spread spectrum technique developed and commercialized by Semtech Corporation [2][3]. It achieves long-range communications and supports ultra low power devices. The LoRaWAN network comprises multiple gateways (GWs), operating in a wide-area, and providing connectivity to the possibly huge amount of deployed end-devices. The LoRa modulation is based on the Chirp Spread Spectrum technique offering low transmission power and robustness from channel degradations. Transceivers in the LoRaWAN GW receive multiple number of messages from the used channels [4]. The spread spectrum provides orthogonal separation between signals by using different spreading factors to the individual signal. Thanks to this, LoRa provides a bidirectional communication at data rates up to 50 kbit/s and is able to cover radio ranges in the order of kilometers [5]. A typical metropolitan LoRaWAN deployment is constituted by multiple-gateways connected to a common Network Server (NetServer). The communication between end-devices and gateways is spread out on different frequency channels and data rates. LoRa uses up to 6 different programmable Spreading Factor (SF): 7, 8, 9, 10, 11, 12. Furthermore, also the adopted bandwidth can be configured: 125 kHz, 250 kHz and 500 kHz (typically 125 kHz for the 868 ISM band.) The selection of the data rate is a trade-off between communication range and message duration, given that communications with different SFs are often assumed not to interfere with each other. LoRa data rates range from 0.3 kbps to 50 kbps. To maximize both battery life of the end-devices and overall network capacity, the LoRa network infrastructure can manage the data rate and RF output for each end-device individually by means of an Adaptive Data Rate (ADR) scheme. The relation between the nominal bit rate and the SF is given as: $R = SF \times \frac{CR}{SF \times BW}$.

In this context, several papers have analyzed the potentialities of LoRaWAN systems in terms of scalability and performance [6][7] and have also highlighted the relevant limits [8]. Specific attention has been recently given to the power and spreading factor allocation in order to avoid the near-far problems by allocating distant users to different channels [9] and to definition of mechanisms to configure the communication parameters of LoRa networks in dense IoT scenario [10]. In the paper [1] the EXPLoRa-Air Time (AT) has been defined. This solution playing with the LoRa modulation SF technique improves the throughput and data extraction rate performance compared whith the ADR [2]. The main result of EXPLoRa-AT is that there is the possibility, in the radio range of a single LoRa Gateway, to allocate different spreading factors to the transmitting end-devices, with the goal of assuring a similar time on air period to all of the them and inducting less collisions and an higher throughput.

In this work we extend the use of the idea of balancing the Air-Time (AT) in a multi-gateway scenario. To this aim we define an ADaptive Mitigation of the Air-time pressure in IORa (AD MAIORA) for a multi-gateway scenario. AD MAIORA is based on the following main considerations: i) there are multiple gateways that may receive the same message by the same end-device if a given SF is used; ii) the use of the same SF by multiple end-devices sharing the same GWs, in the same coverage area, my overload the wireless medium causing collisions. Some SFs may result overloaded and the idea of letting the different SFs having a similar occupation (measured as Time-on-Air) goes in the direction of mitigating message collisions.

Since the final objective is to increase the Data Extraction Rate (DER), i.e., the amount of messages that arrive at least at one GW without corruption, AD MAIORA proposes an heuristic able to keep high the DER by reducing the pressure that some nodes put on some GWs by using a given SF.

II. AD MAIORA - ADAPTIVE MITIGATION OF THE AIR-TIME PRESSURE IN IORA

In a multi-gateway scenario, the assignment of SFs to nodes of the network has a different impact compared to scenarios with a single gateway. In this case, in fact, we have to consider not only the impact that a node transmitting at a given SF has on a single GW, but also the impact it has on the other GWs of the network in its radio visibility. By changing the SF of a node we can give to it more radio visibility but this may also influence the overall network depending on the reciprocal
distances of the node from all the gateways.

The idea of AD MAIORA is to make a SF allocation that allows to distribute the load in a fair way not only on the different SFs but (as far as possible) on the different GWs. AD MAIORA takes into account that we can assign higher SF values to nodes eligible for a lower SF (e.g., SF=8 for nodes at SF=7) while the opposite cannot be applied; furthermore, we start from a basic assignment that depends on sensitivity thresholds (the ones used in the ADR scheme [6]).

To have an example, let’s consider a scenario with two GWs and only two nodes both located in the overlapping area between GWs. Let’s assume that both EDs are positioned in the SF8-range of both GWs: with ADR we will have both nodes with the same SF=8, but we could also increase the SF of one ED in order to not create interferences with the other one. It is important to take into account that moving nodes to an high SF always causes an AT increase since, for high SF values, the duration of the chirp increases too. We want to level off the AT values over all the SFs of each GW and we can assign higher SF to nodes (on a given gateway) only if this will not increase the maximum AT for that gateway.

A. ADR\textsubscript{MGW}: ADR in multiple gateway scenarios

In AD MAIORA we use an ADR version which is compatible with multiple gateway scenario, and we named it \textit{ADR\textsubscript{MGW}}. The algorithm is the same as the ADR for a single-gateway, except that, the SF assigned to a node is the lowest assuring at least the visibility to 1 gateway. In other words, if at SF7 a node is able to reach fewer gateways than at SF8, \textit{ADR\textsubscript{MGW}} assigns SF7.

B. AD MAIORA algorithm

The AD MAIORA algorithm is based on the concept of pressure, i.e., the weight in terms of AT that quantifies the load on each GW for all the SFs. Let us denote by \( N \) the number of end devices (also denoted by nodes) in the considered multi gateway scenario and by \( N_{GW} \) the number of gateways. Let is GW the set of gateways in the considered area. Let us consider \( RSSIT \), a \([N_{GW} \times N]\) matrix storing all the RSSI values measured in the network: the generic element \( RSSIT[i,j] \) represents the RSSI value with which the \( i \)-th GW receives the signal of \( j \)-th ED. Given \( N_{SF} \) possible SFs (7 in the LoRa legacy scheme) and \( N_{BW} \) possible spectrum bands for the LoRa communications, we identity as \( S \) a \([N_{SF} \times N_{BW}]\) matrix collecting the sensitivity thresholds: the generic element \( S[i,j] \) is the threshold for the \( i \)-th SFs and the \( j \)-th BW value. Let us denote \( sf_{\text{cost}} \) as a \([1 \times N_{SF}]\) vector collecting the basic AT values for each SF i.e. \([1.0, 2.0, 3.56, 7.12, 14.23, 24.93] \). In LoRa, the AT (or packet transmission time) is a value depending on multiple parameters such as the SF [7, · · · , 12], the BW [125 kHz, 250 kHz, 500 kHz], the header and payloads lengths, the Code Rate (CR) and also depends on two flag variables i.e. DE (= 1 if low Data Rate Optimization is enabled) and H (= 0 when the header is enabled).

This can be expressed as \( AT = T_{\text{preamb}} + T_{PL} \) that is the sum of the preamble time \( T_{\text{pream}} \) and the payload time \( T_{PL} \): in order to compute these values we need to define the symbol time \( T_{sym} = \frac{2^{SF}}{BW} \) depending on the SF and the BW. So, the duration of preamble is given by \( T_{\text{pream}} = (n_{\text{pream}} + 4.24) \cdot T_{sym} \) where \( n_{\text{pream}} \) is the number of programmed preamble symbols. Instead, to compute \( T_{PL} \) we need to evaluate how many symbols make up the packet payload and header as follows:

\[
PL_{sym} = 8 + \max\left[\frac{8PL - 4SF + 44 - 20H}{4(SF - 2DE)}\right] \cdot (CR + 4), 0]
\]

Once the \( AT \) has been defined, we can define \( sf_{\text{press}} \) as a \([N_{SF} \times N_{GW}]\) matrix collecting the sum of the ATs for each SF and each GW: the generic element \( sf_{\text{press}}[i,j] \) represents the sum of all ATs (in milliseconds) deriving by all nodes transmitting at the \( i \)-th SF value on the \( j \)-th GW.

Our purpose is to reduce and balance the pressure on the different GWs playing with the SFs and the different nodes visibilities to the GWs. To this aim we identify two variables: \( sf_{\text{map}} \) and \( nodetosf \). The former is the current mapping of SFs \([N \times N_{GW}]\), the latter is the final mapping (after AD MAIORA) between nodes and the new SFs. The \( sf_{\text{map}} \) is initialized with \( ADR_{MGW} \). The algorithm is characterized by multiple iterations during which the \( sf_{\text{map}} \) variable is used in order to associate a temporary SF to a node. At the end of each run the value of the \( i \)-th element of \( nodetosf \) is updated with the value computed in \( sf_{\text{map}} \). The algorithm runs until all the changes are added to \( nodetosf \). For each iteration, AD MAIORA performs three steps: first, calculates the \( sf_{\text{press}} \); then, finds the best node than can be moved to another SF in order to improve the overall performance; finally the algorithm identifies an optimal SF value and, if there is “free Air Time” left, both node and new SF are added to \( nodetosf \). In the following subsections [II-C and II-D] we will explain in detail how the best node and the best SF are chosen.

C. Choosing the best node

In order to find the best candidate node for a movement from a SF to another SF, we derive from the pressure table \( sf_{\text{press}} \) the most stressed gateway \( w_{MF} \) and the most overloaded SF for that gateway \( w_{SF} \). Then we can build a set of nodes, named \textit{stressingNodes}, composed by nodes transmitting with \( SF = w_{SF} \) in the coverage area of the GW \( w_{MF} \). Let us introduce a weight \( W_n \) for each node \( n \) belonging to \textit{stressingNodes} that represents the benefit obtained (in terms of AT) if we set a greater SF to the node \( n \). Since we use \( ADR_{MGW} \) to assign the initial SFs values, we can define \( SF^* \) as a set comprising the overall spreading factor values higher than \( w_{SF} \), and we assume \( sf^* \) as an element of the set above.

We compute for each \( i \)-th GW and node \( n \in \textit{stressingNodes}, \Delta_{gw, sf} \) for all values in \( SF^* \) as the difference between the maximum \( \lambda_{gw} \) (defined as in Eq. 2) and the sum of ATs caused by nodes with \( sf^* \) to GW \( gw \) as in Eq. 3:

\[
\lambda_{gw} = \max\{sf_{\text{press}}[sf, gw] | \forall sf \in SF\}
\]

\[
\Delta_{gw}^{(n)} = \{\lambda_{gw} - sf_{\text{press}}[sf^*, gw] | \forall sf^* \in SF^*, sf^* > w_{SF}\}
\]

Hence, for all GW, we compute the minimum \( \text{Min}(\Delta_{gw}^{(n)}) \) among the possible \( \Delta_{gw}^{(n)} \) values we take the smallest for each.

\( ^{\text{1}} \) The \( i \)-th gateway is considered for node \( n \) only if \( n \) is visible to the gateway and the \( RSSIT(gw, n) \geq S[sf^*, bw] \).
Algorithm 1 Choose the best node
1: function bestNode(RSSS, sfmap, sfpress, S, GW, N, bw)
2: wSF, worstGW = Max(sfpress)
3: stressingNodes = \{n | ∀n ∈ N, sfmap[worstGW, n] = wSF\}
4: candidates = \{
5: for n ∈ stressingNodes do
6: for gw ∈ GW do
7: \(\Delta_{gw}^{(n)} = \{
8: for sf^* ∈ [7, 8, 9, 10, 12, 11] and sf^* > wSF do
9: if RSSS[gw, n] ≥ sf[sf^*, bw] then \(\lambda_{gw} = Max(sfpress, gw)\)
10: if \(\lambda_{gw} > sfpress[sf^*, gw]\) then \(\Delta_{gw}^{(n)} = \{\lambda_{gw} - sfpress[sf^*, gw]\}\)
11: end if
12: \end for
13: \end for
14: \(\Delta = \{\\}
15: \)\end for
16: for gw ∈ GW do
17: if RSSS[gw, n] ≥ sf[sf^*, bw] then
18: \(\lambda_{gw} = Max(sfpress, gw)\)
19: \end if
20: end for
21: nextSF = \{\}
22: for sf^* ∈ [7, 8, 9, 10, 12, 11] and sf^* > wSF do
23: end function

Algorithm 2 Find the best spreading factor
1: function bestSF(n, wSF, RSSS, sfpress, sfcost, S, GW, bw)
2: nextSF = 0, nextAT = 0
3: for sf^* ∈ [7, 8, 9, 10, 12, 11] and sf^* > wSF do
4: \(\Delta_{sf^*} = \{\}
5: for gw ∈ GW do
6: if RSSS[gw, n] ≥ sf[sf^*, bw] then \(\lambda_{gw} = Max(sfpress, gw)\)
7: if \(\lambda_{gw} > sfpress[sf^*, gw]\) then \(\Delta_{sf^*} = \{\lambda_{gw} - sfpress[sf^*, gw]\}\)
8: \end if
9: \end for
10: \(\Delta_{sf^*} = \{\\}
11: end if
12: \)\end for
13: \(\{\\}
14: end if
15: \)\end for
16: nextSF = sf^*
17: return nextAT, nextSF
18: end function

GW because, for that node n, it allows to change SF for that node by keeping low the increment of total AT.
We add then this value at the set \(\Delta\) collecting all minimum values for all GWs; finally, for each node n, \(W_n\) is:

\[ W_n = \sum_{i=1}^{N_{GW}} \text{Min}(\Delta_{gw}^{(n)}) \]  

(4)

Essentially, we are looking for the spreading factor with the maximum free space among all considered gateways.

### D. Find the best spreading factor

Once the best node has been found, we need to evaluate which SF value could be the best choice. In order to find the optimal SF, we need to evaluate how changing the SF impacts on the network (the calculus is similar to the one considered in II-C, except that now we are subtracting also \(\text{sfcost}[sf^*]\) as in Eq. \[5\]). Hence, the heuristic calculates \(\Delta_{sf^*}\), collecting all the differences between \(\lambda_{gw}\), the pressure on the current GW at \(sf^*\) and the cost of that node at \(sf^*\) in terms of Air Time; this is stored in the set \(\Delta_{sf^*}\), computed as in Eq. \[5\] for all SF\(^*\) values of the node selected in II-C. The best SF is the one that allows to get the highest \(\text{Min}(\Delta_{sf^*})\) as in \[5\].

\[ \Delta_{sf^*} = \{\lambda_{gw} - sfpress[sf^*, gw] - sfcost[sf^*] | \forall gw \in GW\} \]  

(5)

\[ \text{nextAT} = \text{Max}\{\text{Min}(\Delta_{sf^*}) | \forall sf^* \in SF^*\} \]  

(6)

The performance evaluation of AD MAIORA has been carried out by extending LoRaSim simulator [6]. We considered different scenarios where N End-Devices (EDs) are randomly distributed in a bi-dimensional space around one or more GWs. We analyzed two different topologies for the node’s location as shown in Figure [1] a balanced one in Figure [1](a) where the 60% of nodes is located in a central area between the gateways, and an unbalanced one, in Figure [1](b) where 60% of nodes are located around a specific gateway in a range of 50 meters. All simulation results are represented with their 95% confidence interval. The nodes work using the communication transmission parameters reported in Table [I].

Regarding the duty cycle, in Europe its values are established by ETSI EN 300.220 standard [11], which defines different values for different sub-bands; in particular, one of those sub-bands \(G3: 869.4 - 869.65MHz\) has a duty cycle of 10%. This means that considering a message period equal to 10 seconds, at the maximum air time value (quantified at about 1 second for SF12) there are 9 seconds of waiting time before a new retransmission [3] [4]. Since in the simulations we considered a message period equal to 10 seconds, we are exactly in the case where the duty cycle is at 10% (not higher).

The goal of our simulations is to prove the effectiveness of the algorithms at a high pressure level: using the G3 sub-band we get a simulation where the devices transmit much more frequently (so with greater probability of collision). The stress level in our simulations is therefore the maximum possible, while remaining within the limits imposed by LoRaWAN and the standard for the ISM band. Furthermore, in our simulations it has been proposed an allocation algorithm called "probabilistic ADR" that distributes the SFs following a probability distribution inversely proportional to the air time. In this way, considering only the number of nodes for each SF value (and not their position relatively to GWs), we will get a term of comparison of AD MAIORA against a simpler approach.
As a first evaluation, in Figure 2 we plot the DER and throughput as a function of the message period with six different lines, each one representing different scenarios and allocations. The scenario is characterized by 4 GWs and 500 EDs, located according to the two topologies considered above. Each ED forwards a message with a decreasing rate. We note that in both scenarios the ADR attains the highest throughput only for high message periods (low message rates) while AD MAIORA presents improvements at high network loads (i.e. high message rates). The figure shows a gray area representing the values of message period violating the possible duty cycle. In order to clearly plot the different scenarios, we used continuous lines for balanced scenarios and dashed lines for unbalanced ones. In case of balanced scenarios, with AD MAIORA there is a DER improvement, especially in case of lower message periods; in the other case we have for probabilistic ADR both good DER and throughput because the assignment made with AD MAIORA (when the 60% of nodes are positioned around a single GW) is similar to EXPLoRA-T: this is due to the fact that in both cases the number of EDs with the same SF value is inversely proportional to the air time. Analyzing the plot in Figure 2(b) we can derive that, for a message period greater than 100 seconds, the ADR has a good throughput so it would not be possible to bring improvements. On the contrary, we can improve the performance in case of highly stressed scenarios, e.g. when nodes transmit with a message period equal to 10 seconds.

In Figure 3 we evaluate the performance in terms of DER and throughput as a function of the number of nodes. In this case, each device transmits with a message period equal to 10 seconds, and nodes are placed as in Figure 1(a). As expected, by increasing the number of nodes, both the DER and the throughput decrease following a monotonous trend, but it is
possible to verify that with AD MAIORA the performance is better than ADR of a nearly constant value, until the devices do not exceed a certain number. In fact, we can observe that the performance drops significantly when the gateways are subjected to a great pressure, that is if the nodes start to become more than 250. For this reason, in the following experiments we analyze scenarios with a number of nodes equal to 500, in order to evaluate the performance of AD MAIORA in critical contexts highlighting its ability to make the best use of the GWs present in the scenario.

We also plotted the performance according to the GWs number considering a scenario with fixed nodes distributed as the above topologies. In case of single-GW, 500 EDs are distributed in a circular range of 50 meters around the GW. The DER and throughput behaviors are shown in Figure 4 for two different spreading factor allocations: i) $AD_{MGV}$ and ii) AD MAIORA, where both are considered for the balanced (Fig. 1(a)) (continuous lines) and unbalanced (Fig. 1(b)) (dashed lines). The message period is set to 10 seconds. We notice that AD MAIORA attains a good performance gain as the number of GWs increases. In particular, we can evaluate that ADR in both scenarios (red and orange lines) has an unusual trend between values 1 and 4 in abscissa, so passing from a single-gateway scenario to a 4-gateways one. This irregular trend can be explained by observing Fig. 6 which shows how many nodes have the same SF value when
the number of gateways increases in the topology: we can infer that the gateways increment leads to a greater radio coverage, on the other hand we will have a greater number of nodes having the same SF value due to the overlapping areas. Indeed, we can notice in Fig. 6(a) that in case of 8 GWs and ADR allocation, almost all nodes have the same $SF = 7$ value, instead Fig. 6(b) shows that the number of nodes assigned for each SF value is inversely proportional to the respective air time, like EXPLoRa-AT in [1]. Furthermore, in the same figure, we show the performance evaluated in case of Fig. 1(b) plotted with dashed lines. We can notice a more continuous trend of the ADR curve (yellow) and a slightly higher performance compared to the balanced topology case; we can explain this behavior by considering the partial DER for each gateway: as shown in Fig. 5 in case of balanced topology (red and blue bars), each gateway has approximately the same DER value because the majority of nodes are in the overlap central areas and so, for instance, a node having $SF = 9$ located in the middle area overloads the channel of all gateways. Instead, in case of unbalanced topology (orange and green bars), there is also an unbalanced DER for each gateway; in particular, there is a gateway with a very low DER corresponding to the overloaded one, but others have a good DER due to a low traffic load.

V. CONCLUSIONS

In the IoT market, the Low Power, Low Range technologies (LoRa) are emerging. To fully exploit the LoRa potentials it is critical to support hundreds of devices also in stressed configurations (e.g. multiple gateways interconnected to nodes transmitting with a 10% of duty cycle and a high message rate). To this aim, in this paper, we presented AD MAIORA, an algorithm that adaptively allocates the spreading factor to nodes in the radio visibility of multiple gateway with the aim of reducing the pressure on the gateways of nodes using the same SFs. We showed, via simulations, that AD MAIORA, by suitably allocating the SFs attains a noticeable gain with respect to the classic ADR approach. This result is particularly evident in case of very loaded and unbalanced scenarios.

REFERENCES

[1] 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 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). IEEE, 2017, pp. 1–8.
[2] Semtech, “LoRa,” http://www.semtech.com/, 2015.
[3] M. Bor, J. Vidler, and U. Roedig, “LoRa for the Internet of Things,” in Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks, ser. EWSN ’16, 2016, pp. 361–366.
[4] N. Sornin, M. Luis, T. Eirich, T. Krampe, and O. Hersent, “LoRaWAN Specification,” https://www.lora-alliance.org/ 2015.
[5] J. Petäjäjärvi, K. Mikhaylov, M. Pettissalo, J. Janhunen, and J. Iinatti, “Performance of a low-power wide-area network based on lora technology: Doppler robustness, scalability, and coverage,” Int. Journal of Distributed Sensor Networks, vol. 13, no. 3, pp. 1–16, 2017.
[6] M. C. Bor, U. Roedig, T. Voigt, and J. M. Alonso, “Do LoRa Low-Power Wide-Area Networks Scale?” in 19th ACM Int. Conf. on Modeling, Analysis and Simulation of Wireless and Mobile Systems, ser. MSWiM ’16, 2016, pp. 59–67.
[7] T. Voigt, M. Bor, U. Roedig, and J. Alonso, “Mitigating inter-network interference in lora networks,” in Proceedings of the 2017 Int. Conf. on Embedded Wireless Systems and Networks, ser. EWSN &#8217;17, 2017, pp. 323–328.
[8] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Sequi, and T. Watteyne, “Understanding the Limits of LoRaWAN,” IEEE Communications Magazine, vol. 55, no. 9, pp. 34–40, 2017.
[9] B. Reynolds, W. Meert, and S. Pollin, “Power and spreading factor control in low power wide area networks,” in 2017 IEEE International Conference on Communications (ICC), May 2017, pp. 1–6.
[10] M. Slabicki, G. Premsankar, and M. D. Francesco, “Adaptive Configuration of LoRa Networks for Dense IoT Deployments,” in 16th IEEE/IFIP Network Operations and Management Symposium (NOMS 2018), 2018, pp. 1–9.
[11] ETSI, “300 220-1 v2. 4.1 (2012-01),” Electromagnetic compatibility and Radio spectrum Matters (ERM), 2012.