Multi-Agent Q-Learning Algorithm for Dynamic Power and Rate Allocation in LoRa Networks

Yi Yu, Lina Mroueh, Shuo Li, Michel Terré

Multi-Agent Q-Learning Algorithm for Dynamic Power and Rate Allocation in LoRa Networks

Yi Yu[(1,2)], Lina Mroueh[(1)], Shuo Li[(1)] and Michel Terré[(2)]
(1) Institut Supérieur d'Electronique de Paris, 75006 Paris, France
(2) Conservatoire national des arts et métiers, 75003 Paris, France
Email: yi.yu@isep.fr, lina.mroueh@isep.fr, shuo.li@isep.fr, michel.terre@cnam.fr

***Abstract*—In this paper, we consider a Low Power Wide Area Network (LPWAN) operating in a licensed-exempt band. The LoRa network provides long-range, wide-area communications for a large amount of objects with limited power consumption. In terms of link budget, nodes that are far from the collector suffer collisions caused by nodes that are close to the collector during the data transmissions. Chirp Spread Spectrum (CSS) modulation is adopted by assigning different Spreading Factors (SF) to active sensors to help reduce destructive collisions in LoRa network. In order to improve the energy efficiency and communication reliability, we propose an application of multi-agent Q-learning algorithm in the dynamic allocation of power and SF to the active nodes for uplink communications in LoRa. The main objective of this paper is to reduce power consumption of the uplink transmissions and to improve the network reliability.**

***Index Terms*—LPWAN, LoRa, Internet of Things (IoT), resource allocation, machine learning, Reinforcement Learning, multi-agent Q-learning, energy efficiency.**

I. Introduction

Reinforcement Learning (RL) is a branch of Machine Learning (ML) techniques which is of great interest because of its large number of practical applications in control engineering, multi agent systems and operation research. It is an efficient and dynamic decision-making tool.

Thanks to the environment-agent interaction scheme in RL, it enables a system to be controlled in such a way as to maximize a numerical measure of performance that expresses a long-term objective [1]. It shows great potential for system controls in real engineering cases.

In particular, we are interested in this paper in the Q-learning which is a popular RL algorithm. It adopts off-policy Temporal-Difference (TD) methods for agents to learn how to act optimally to control problems. Considering multi-agent systems in a variety of fields, including robotics, distributed control, and telecommunications, multi agent Q-learning algorithm gives outstanding performance in decision-making problems for these complicated scenarios [2]–[5]. Based on it, we propose in this paper to apply it to the SF and power allocation problems in LPWAN.

LPWAN is well suited for massive IoT network deployment as it can provide long-range, wide-area, low-power consumption communications for thousands of connected devices. There are different types of low power wide area communication technologies on the market, i.e unlicensed LoRa [6]–[8] and Sigfox [9] technologies and licensed NB-IoT technology [10], [11], etc. In this paper, we focus on LoRa network.

LoRa technology uses Chirp Spread Spectrum modulation which allocates spreading factors to each active sensor node to mitigate destructive collisions in the overall network. Higher spreading factors can protect the devices from the interference caused by signals transmitted simultaneously from devices closer to the collector. To improve network reliability, we need to assign different spreading factors in an adaptive way to the active sensors.

In LPWAN, energy efficiency always attracts a lot of attention. Normally, a single IoT module is expected to serve for around ten years with very low energy consumption. By reducing power consumption during each message transmission for uplink communications in LPWAN, it can improve the lifetime of IoT devices and offer remarkable financial benefits. In this paper, we fully consider the energy efficiency by select distinct transmit power values for each sensor node for uplink communications to achieve a better power control solution for LoRa network.

Some relative work has been done by using different methods. For example, in [12], a power, channel and spreading factor selection algorithm is proposed to avoid near-far problem and decrease the packet error rate in LoRa network. A distributed learning approach is introduced for self-organized LoRa networks in [13]. Inspired by them, we use multi-agent Q-learning approach to solve the concerned power and SF allocation problems dynamically.

The rest of the paper is organized as following. An unlicensed LoRa network model is defined in section II. Then, we review in Section III the multi-agent Q-learning scheme. Next, in section IV, multi-agent Q-learning algorithm is applied to solve the problems of SF selection and power allocation in LoRa network. In section V, the performance evaluation in resource allocation with respect to network reliability and power efficiency is presented. Finally, section VI concludes the paper.

II. LoRa Network model

In this section, we consider a single unlicensed LoRa network cell. The cell coverage radius is R. Assume the collector is located in the cell center which is marked as origin o. Number of N active sensor nodes are randomly

distributed in the given cell. Let $|x|$ denotes the distance between the active sensor node x and the collector o. The out-of-cell interference is ignored in our case. Considering the uplink transmission of LoRa network, each active sensor has 5 different choices for the transmission power P_t: 2, 5, 8, 11 and 14 dBm [12]. The sensor nodes and the collector have omnidirectional antennas with 0 dBi of antenna gains. For the sensor node x, the received power P_r at the collector is:

$$P_r(x) = P_t \alpha |x|^{-\beta} A_f, \tag{1}$$

where α and β are respectively the attenuation factor and the path-loss exponent that are computed from Okumura-Hata model. A_f is the random fading coefficient with Rayleigh distribution.

The LoRa network adopts CSS modulation, which is a spread-spectrum technology [7], [14], [15]. According to [16], CSS modulation has several key properties, it offers high robustness and resists the Doppler effect while having a low latency.

In LoRa network, the active sensors located far away from the collector suffer the collision caused by sensors that are closer to the collector. This CSS spreading technique can protect the cell edge sensor nodes from the nodes in the proximity of the network collector. It features 6 possible spreading factors (SF = 7 to 12) to the active sensors according to the receiver sensitivity and hence by the threshold communication ranges. Table I shows the corresponding SINR threshold ranges according to the values of SF with the sub-bandwidth equal to 125 kHz. With CSS modulation, each symbol transmits SF bits, has a time duration T and occupies a bandwidth B, we have

$$2^{\mathrm{SF}} = T \times B \tag{2}$$

For the same sub-bandwidth, the high spreading factor transmits longer time on air which means the communication distance increases. Concern the data rate R_b (bits/s),

$$R_b = SF \times \frac{B}{2^{SF}} \frac{4}{(4+CR)} \tag{3}$$

with CR being the code rate. A high spreading factor better prevents transmission errors, but at the cost of a reduced data rate. LoRa network uses high spreading factors for the weak signal or the signal suffering high interference.

TABLE I
SINR THRESHOLD γ_{SF} WITH SUB-BANDWIDTH $B = 125$ KHZ

SF	7	8	9	10	11	12
γ_{SF} (dB)	-7.5	-10	-12.5	-15	-18	-21

In this paper, our objective is to manage the resource allocation in a dynamic manner. Transmission power and SF are allocated to each active sensor node to ensure the overall network communications. We aim at increasing the communication reliability for each sensor while keeping the energy consumption as low as possible. In LoRa network, the access to the shared medium is managed by Aloha protocol. The transmitted signals on the same sub-medium interfere with each other. The collector receives in addition to its intended attenuated signal.

Assume that Φ_i denotes the set of interfering nodes. For an intended signal sent by node x, the interferer $y \in \Phi_i$. The power of the interference is weighted by a correlation factor denoted by $c(x, y)$. The expression of the interference is then,

$$I_x \approx \sum_{y \in \Phi_i} c(x,y) \alpha |y|^{-\beta} A_f P_t(y), \tag{4}$$

where $P_t(y)$ is the transmit power of interferer y. The inter-correlation factor $c(x, y)$ is calculated in [17]. Table II presents the $c(x, y)$ values according to different spreading factors of the transmitter x and the receiver y with the sub-bandwidth of 125 kHz.

The received SINR for the sensor node x at the given collector o is calculated as follows,

$$\mathrm{SINR}_x = \frac{P_t \alpha \, |x|^{-\beta} \, A_f}{N_0 + I_x}, \tag{5}$$

with $N_0 = KTB$ being the additive thermal noise. K is the Boltzmann constant, T is the noise temperature and B is the bandwidth.

III. MULTI-AGENT Q-LEARNING ALGORITHM

Reinforcement Learning has achieved many successes in decision-making systems. Figure 1 demonstrates the general RL framework. The environment and the agent can have interaction and learn from it to make decisions. S is the set of possible environment states. At time t, observing the environment, a state $s \in S$ is observed and passed to the agent. Then, according to the policy π, the agent decides to take action $a \in A$, where A is the set of actions available for state s. As a is performed, the agent earns a reward $r(s, a)$ and the environment turns to a new state s'.

TABLE II
THE INTER-CORRELATION FACTOR $c(x, y)$ WITH SUB-BANDWIDTH $B = 125$ KHZ [17]

SF	7	8	9	10	11	12
7	0	16.67	18.20	18.62	18.70	18.65
8	24.08	0	19.70	21.27	21.75	21.82
9	27.09	27.09	0	22.71	24.34	24.85
10	30.10	30.10	30.10	0	25.73	27.38
11	33.11	33.11	33.11	33.11	0	28.73
12	36.12	36.12	36.12	36.12	36.12	0

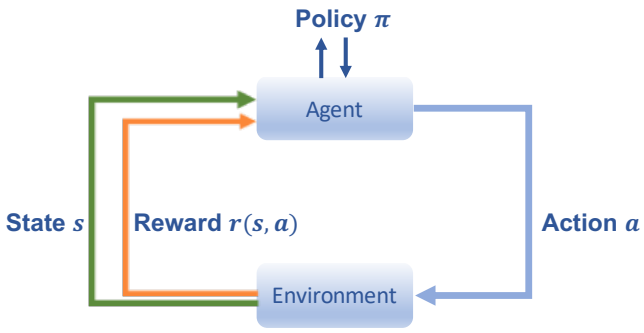

Fig. 1. General Reinforcement Learning Framework

Since the agent determines a policy π, the choose of action a in a given state s can be described as $\pi(s) = a$. To evaluate the performance of policy π, the state-action value function $Q(s, a)$ is introduced as follows:

$$Q(s, a) = \mathbb{E}\left[G_t | s_t = s, a_t = a\right], \tag{6}$$

where G_t is the discounted future cumulative reward,

$$G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots + \gamma^{n-t} r_n, \tag{7}$$

with γ as the discount factor which is a constant and $\gamma \in [0, 1]$.

The objective of reinforcement learning is to find the best policy $\pi^\star$ that maximizes the state-action value function which can be described as:

$$\pi^\star(s) = \arg\max_{a \in A} Q(s, a). \tag{8}$$

A. Q-learning Function

Q-learning uses a non-deterministic policy, i.e. a function mapping each state to a set of actions so that the agent can choose one among them, while it is based on a sampling of other policies instead of the current policy alone. The Q-learning process keeps an estimate and an update of the Q-function, it can be written as follows:

$$\begin{aligned} Q(s, a) &= Q(s, a) \\ &+ \sigma \left\{ r(s, a) + \gamma \max_{a \in A} Q\left(s', a'\right) - Q(s, a) \right\}, \end{aligned} \tag{9}$$

where $\sigma \in [0, 1]$ is a learning rate.

B. Multi-Agent Q-Learning Algorithm

Assume that there are n agents in the system. At a given time t, action $\boldsymbol{a} = \{a_1, a_2, ..., a_n\}$ is executed by all the agents with $a_i \in A$ being the action chosen by the i-th agent. Observing the environment, a current state $s \in S$ and a total reward $r(s, \boldsymbol{a})$ are obtained and passed to all the agents. Based on state s and reward r, each agent updates its own Q-learning function and chooses a new action $a_i{}'$ for next operation separately. Then, the action $\boldsymbol{a}' = \{a_1', a_2', ..., a_n'\}$ is executed. The workflow of multi-agent Q-learning keeps the same as the single-agent Q-learning algorithm. But the state and reward information are distributed to all the agents and then a decision is made by each agent independently.

IV. SF Allocation and Power Control with Multi-Agent Q-Learning Algorithm

The framework of multi agent Q-learning is presented in section III. In this section, we try to solve the problems of SF allocation and power control in a licensed-exempted LoRa network with multi-agent Q-learning algorithm. According to the network model mentioned in section II, for each active sensor node connected to the collector o, it must be assigned one SF value among 6 possible spreading factors $[7, 8, 9, 10, 11, 12]$ and one transmit power value out of 5 achievable transmit power values including $[2, 5, 8, 11, 14]$ dBm.

Algorithm 1 Multi-agent Q-learning algorithm for SF allocation and Power Control in LoRa Network

Input: Positions of N active sensor nodes

Output: SF and transmit power values for each active sensors

Sort N sensor nodes by distance to the collector from near to far;
Set $n = N$ agents and generate n Q-learning function $Q_i(S, A), i \in [1, n]$;
Initialize each Q-learning function $Q_i(S, A), i \in [1, n]$ with random values;
for episode $= 1, M$ **do**
  Set the initial observation state s.
  for $t = 1$: until s is a terminal state **do**
    For each agent:
    with the probability ϵ select a random action $a_i = (SF, P_t)$,
    otherwise select $a_i = \max_a Q_i(s, a_i)$;
    Execute action $\boldsymbol{a} = \{a_1, a_2, ..., a_n\}$ in emulator and observe reward r and state s';
    If s' is a terminal state, set the value function target y_i to $r(s, \boldsymbol{a})$;
    Otherwise set it to: $y_i = r(s, \boldsymbol{a}) + \gamma \max_a Q_i(s', a_i)$;
    Compute the critic parameter update $\Delta Q_i = y_i - Q_i(s, a_i)$;
    Update the critic using the learning rate σ: $Q_i(s, a_i) = Q_i(s, a_i) + \sigma * \Delta Q_i$;
    Set the observation state s to s'
  end for
end for
return Results

Following the multi-agent Q-learning algorithm, the LoRa network model is the environment that provides the state s and reward r information for the agents. We set $n = N$ agents for the system. Each agent corresponding to an active sensor node. A single agent here is a Q-learning function which selects an action for an active node. The action is a

pair of SF and P_t values with probability ϵ or selects an action with probability $(1-\epsilon)$ by maximizing the Q value,

$$a_i = \arg\max_a Q_i\left(s, a_i\right).$$

After the execution of the action $\boldsymbol{a_i}$, the agents obtain the new state information $s \rightarrow s'$ and a reward r. Considering the network reliability and the power efficiency, the reward r_i for the i-th sensor node is calculated as follows:

$$r_i = \delta_i(t) \cdot \varphi + \delta_i(t) \cdot (1-\varphi) \cdot \frac{P_{t\,max}}{P_t}. \tag{10}$$

where φ is a design parameter offering a tradeoff between the network reliability and energy efficiency. P_t is the transmit power for the i-th sensor node. $\delta_i(t) \in \{0, 1\}$ which indicates whether the i-th sensor has a stable connection to the collector o or not. With a given SF value SF_i and a transmit power P_t for the i-th sensor node, we can calculate the SINR_i on the collector side for each sensor node. In table I, the SINR threshold γ_{SF} for different SF values is given in detail. Hence, if $\mathrm{SINR}_i \geqslant \gamma_{SF_i}$, the i-th node can successfully send the message to the collector. Otherwise, the transmission fails.

$$\delta_i(t) = \begin{cases} 1 & \mathrm{SINR}_i \geqslant \gamma_{SF_i} \\ 0 & \mathrm{SINR}_i < \gamma_{SF_i} \end{cases} \tag{11}$$

The state of the environment after the execution of the action $\boldsymbol{a}$ is the decimal value corresponding to a binary number consists of all the δ_i. For example, if the number of agents n is equal to 5. There are $2^5 = 32$ states ranging from 0 to 31. At time t, if $\{\delta_1\delta_2...\delta_5\} = \{10110\}$, then the state equals 22 which is the decimal value for this binary number. Obviously, the terminal state is when all the $\delta_i = 1$. In case $n = 5$, the terminal state $s = 31$.

The total reward for the action $\boldsymbol{a}$ is then calculated as follows:

$$r(s, \boldsymbol{a}) = \frac{1}{N}\sum_{i=1}^{N} r_i. \tag{12}$$

Algorithm 1 demonstrates the selection of SF and transmit power for the active sensors in LoRa with multi-agent Q-learning algorithm.

V. Numerical Results

We consider a single LoRa network cell with coverage radius of $R = 10$ km. Assume that N sensor nodes are activated simultaneously with random positions. The sensor nodes and the collector have omnidirectional antennas. Nodes transmitting in the same frequency band generate additive interference with power. The power of the additive thermal noise is $N_0 = KTB$ with $K = 1.379 \times 10^{-23}$ W Hz^{-1} K^{-1}, $T = 290$ K and $B = 125$ kHz. The system parameters are shown in detail in Table III.

We assume without loss of generality a Q-learning algorithm with 5 agents. Set the learning rate $\sigma = 0.9$. Following Algorithm 1, we can obtain the expected SF and transmit

TABLE III
Network Parameters

Parameter	Value
Carrier frequency	868 MHz
Sub-bandwidth	$B = 125$ kHz
Coverage radius	$R = 10$ km
Transmit power	$[2, 5, 8, 11, 14]$ dBm
Spreading factor	$[7, 8, 9, 10, 11, 12]$
Collector height	30 m
Device average height	1 m
Antenna gain	0 dBi
Urban path-loss model	$\alpha = 10^{-10.07}, \beta = 3.52$

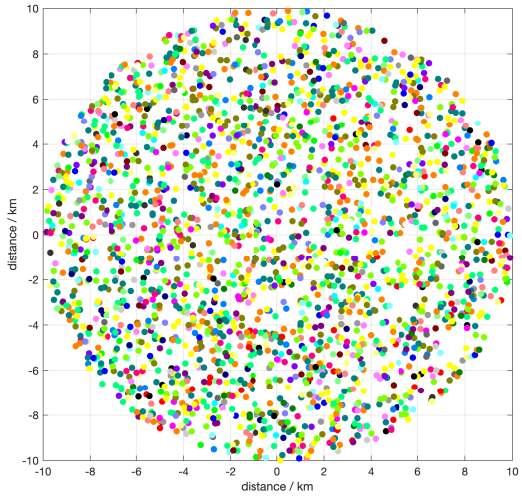

Fig. 2. The distribution of active sensor nodes in a single cell with different choices of SF and transmit power

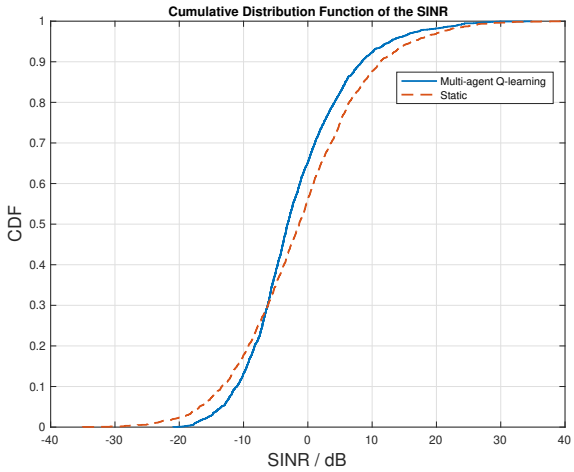

Fig. 3. The Cumulative distribution function of the SINR

power values for all 5 nodes after the training. Figure 2 shows the distribution of active nodes for 500 times of simulation with 5 agents. There are 30 different colors which represent 30 pairs of SF and P_t values. We can notice that the joint allocated SF and transmission power is made in a dynamic way depending on the instantaneous network configuration.

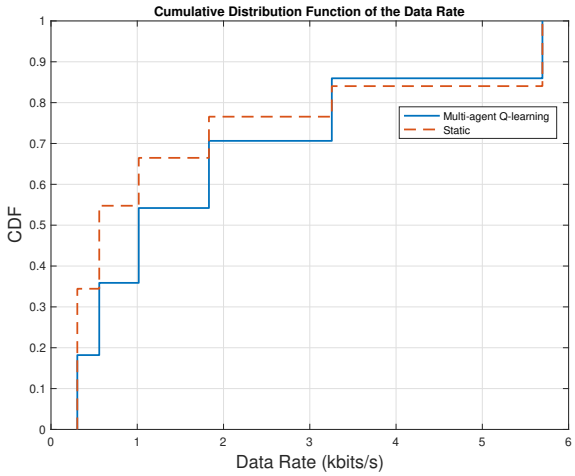

Fig. 4. The Cumulative distribution function of the Data Rate

To make a comparison with our multi-agent Q-learning algorithm, we introduce a static allocation algorithm in which the whole cell is divided into 30 rings. Within each ring, a static SF and P_t is attributed to the active nodes depending on its location. Low SF and low power are attributed to the nodes in the proximity of the collector.

Figure 3 illustrates the Cumulative Distribution Function (CDF) of the SINR. Multi-agent Q-learning algorithm outperforms the static one which means the former has higher reliability then the latter. Meanwhile, it also achieves higher data rate during the transmission as shown in Figure 4.

Figure 5 presents the comparison of the CDF of transmit power for all sensor nodes based on two different algorithms.

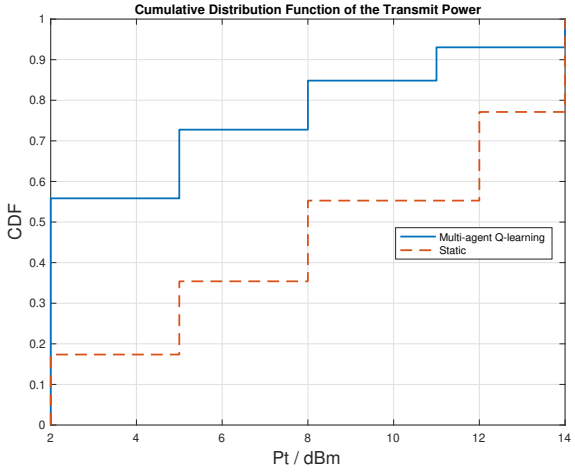

Fig. 5. The Cumulative distribution function of the Transmit Power

Our multi-agent Q-learning algorithm shows large improvements in the energy efficiency. Using the multi-agent Q-learning algorithm, the average value of the transmit power for all active sensors is equal to 4.81 dBm while for the static algorithm it is 8.66 dBm.

VI. Conclusion

In this paper, we have considered the unlicensed LPWAN LoRa network with random Aloha access to the network. We have proposed a multi-agent Q-learning algorithm to jointly allocate a SF and power in the uplink of LoRa network. Each agent has interactions with the environment and based on that, it updates dynamically the policy. We have also evaluated its reliability and energy efficiency performance and compared it with a static allocation algorithm. The simulation results show the advantages of our Q-learning algorithm with respect to SINR, data rate and transmit power.

References

[1] C. Szepesvári, "Algorithms for reinforcement learning," *Synthesis lectures on artificial intelligence and machine learning*, vol. 4, no. 1, pp. 1–103, 2010.

[2] L. Bu, R. Babu, B. De Schutter *et al.*, "A comprehensive survey of multiagent reinforcement learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 2, pp. 156–172, 2008.

[3] M. Abdoos, N. Mozayani, and A. L. Bazzan, "Traffic light control in non-stationary environments based on multi agent q-learning," in *2011 14th International IEEE conference on intelligent transportation systems (ITSC)*. IEEE, 2011, pp. 1580–1585.

[4] E. Rodrigues Gomes and R. Kowalczyk, "Dynamic analysis of multiagent q-learning with ε-greedy exploration," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 369–376.

[5] M. Kaisers and K. Tuyls, "Frequency adjusted multi-agent q-learning," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, 2010, pp. 309–316.

[6] 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*, Oct. 2015. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01231221

[7] "Lora official website," https://lora-alliance.org/.

[8] N. Sornin, M. Luis, T. Eirich, T. Kramp, and O. Hersent, "Lorawan specification," *tECH; REP; LoRa Alliance*, 2016.

[9] "Sigfox official website," www.sigfox.com.

[10] G. T. . v13.1.0, "Cellular system support for ultra low complexity and low throughput internet of things," *[Online]*, November 2015.

[11] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, and A. Ghosh, "NB-IoT System for M2M Communication," *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2016.

[12] B. Reynders, W. Meert, and S. Pollin, "Power and spreading factor control in low power wide area networks," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[13] A. Azari and C. Cavdar, "Self-organized low-power iot networks: A distributed learning approach," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–7.

[14] "Semtech official website," https://www.semtech.com/lora/what-is-lora.

[15] L. Vangelista, "Frequency shift chirp modulation: The lora modulation," *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1818–1821, 2017.

[16] "IEEE 802.15 working group for wireless specialty networks (WSN) Website / IEEE 802.15 documents," http://www.ieee802.org/15//.

[17] Y. Yi, L. Mroueh, D. Duchemin, C. Goursaud, J.-M. Gorce, and M. Terre, "Adaptive multi sub-bands allocation in lora networks," *submitted to IEEE Access*, 2020.