LoRa-MAB: Toward an Intelligent Resource Allocation Approach for LoRaWAN

Duc-Tuyen Ta[1,4], Kinda Khawam[2], Samer Lahoud[3], Cédric Adjih[4], and Steven Martin[1]
[1]ROCS, LRI, Université Paris Saclay, 91190 Gif-sur-Yvette, France
[2] Université de Versailles Saint-Quentin-en-Yvelines, 78000 Versailles, France
[3] Faculté d'ingénierie ESIB, Université Saint-Joseph de Beyrouth, Lebanon
[4] Inria, CRI Saclay-Île-de-France, 91120 Palaiseau, France

***Abstract*—For a seamless deployment of the Internet of Things (IoT), self-managing solutions are needed to overcome the challenges of IoT, including massively dense networks and careful management of constrained resources in terms of calculation, memory, and battery. Leveraging on artificial intelligence will enable IoT devices to operate autonomously by using inherently distributed learning techniques. Fully distributed resource management will free devices from draining their limited energy by constantly communicating with a centralized controller. The present work is devoted to a specific IoT context, that of LoRaWAN, where devices communicate with the access network via ALOHA-type access and spread spectrum technology. Concurrent transmissions on different spreading factors increase the network capacity. However, the bottleneck is inevitable with the expected massive deployment of LoRa devices. To address this issue, we resort to the popular EXP3 (Exponential Weights for Exploration and Exploitation) algorithm to steer autonomously the decision of LoRa devices towards the least solicited spreading factors. Furthermore, the spreading factor selection is cast as a proportional fair optimization problem used as a benchmark for the learning-based algorithm. Extensive simulations were run in a realistic environment taking into account physical phenomena in LoRaWAN such as the capture effect and inter-spreading factor collision, as well as non-uniform device distribution. In such a realistic setting, we evaluate the performances of the EXP3.S algorithm, an efficient variant of the EXP3 algorithm, and show its relevance against the fair centralized solution and basic heuristics.**

***Index Terms*—LoRAWAN, reinforcement learning, EXP3, Spreading Factor selection.**

I. Introduction

The long-term goal of the Internet of Things is to provide low-cost, large-scale, and ultra-durable connectivity for every object that can benefit from being connected. LoRaWAN [1], [2] is a well-known IoT solution over the unlicensed band with a simplified connectivity procedure. It is designed to allow low-powered devices to communicate with the access network over long-range wireless connections. Transmission is possible on one of the 8 channels (frequency plans in Europe [1]) and with one of the 6 available spreading factors. A collision will only occur when two or more devices select the same channel and spreading factor (SF) [3]. However, the latter is inevitable due to the use of random access, the shortage in radio resources and the expected massive deployment of LoRa devices. Accordingly, astute resource management is vital to increase the capacity of LoRaWAN. However, only resource allocation schemes that reduce drastically signaling with the access network in order to offer ultra-long battery lifetimes to LoRa devices are feasible. Therefore, each device must be able to select adequate spreading factors autonomously.

In addition, even with different SFs, a collision between signals on the same channel can occur due to the imperfect orthogonality of SFs, called inter-Spreading Factor collision [4], [5]. Fortunately, if there are concurrent transmissions on the same resource (the same SF and channel), the gateway (GW) is able to successfully receive one of them if its Signal-to-Interference-and-Noise-Ratio (SINR) is higher than a threshold of 6 dB, for any SF. The latter is deemed capture effect [2], [3]. In this paper, we will assess the impact of both phenomena, namely the capture effect (CE) and inter-SF collision, that were overlooked in the literature.

Recent work on distributed selection of radio resources in LoRaWAN had recourse to the Multi-Armed Bandit (MAB) problem [6], [7]. Each end-device is considered as an intelligent agent that chooses a given SF and/or channel to minimize its cumulative regret in comparison with the best fixed allocation that renders the highest reward.

In general, there are two broad MAB models: stochastic and non-stochastic [8]. For stochastic MAB, the reward of each strategy is drawn according to a given probability density function (PDF). Conversely, for non-stochastic MAB, no statistical assumptions are made about the generation of rewards. In particular, adversarial MAB is a non-stochastic MAB where rewards are chosen by an adversary. This formulation can model any form of non-stationarity and is hence adequate for the problem at hand.

In [6], the authors assumed that all end-devices use the same spreading factor and adopted the stochastic MAB algorithm to determine the channel selection. However, such an assumption is impractical in LoRa network due to the mutual coupling between multiple intelligent end-devices. The work in [7] has explored adversarial MAB for resource allocation in an IoT network. However, the capture effect and inter-SF interference were not taken into consideration. More importantly, only uniform device distribution is considered. In this work, we assess the impact of realistic non-uniform device distribution where smart distributed resource allocation becomes crucial.

In this paper, as the distributed selection of the least congested SFs by uncoordinated devices is appropriately modeled by the adversarial MAB problem, we resort to the popular EXP3 (Exponential Weights for Exploration and Exploitation) algorithm [9]. The goal of EXP3 is to steer autonomously the decision of each LoRa end-device towards the least solicited SF while ensuring reactivity to the possible changes that can occur in the common resource usage. In particular, we use the EXP3.S [10], a computationally efficient version of EXP3. We show that the reinforcement learning approach is much more efficient in minimizing the number of collisions, as well as improving the throughput of LoRa network, in comparison with a uniform distribution or a trivial random distribution over the set of SFs. More importantly, we define a proportional fair optimal problem for the SF selection as a benchmark for the EXP3.S algorithm and show that the latter displays little discrepancy with the optimal problem.

The rest of this paper is organized as follows. The system model is presented in Section II. The optimal formulation for the spreading factor selection is casted in Section III. In Section IV, a distributed learning-based approach is investigated to minimize the number of collisions for LoRa end-devices, magnified by inter-SF collision and dense deployment. Performances of the proposed approach are evaluated in Section V. Conclusion is given in Section VI.

II. The System Model

In this paper, we consider a LoRaWAN-type network composed of one gateway located at the center of a disc-shaped network of radius R, and N end-devices. Communications in LoRaWAN occur in one of the 8 channels in the public ISM band; each channel has a bandwidth of 125 KHz in Europe (see [1]). High resiliency to noise and interference is essential to operate efficiently in the ISM band. To this end, the chirp spread spectrum modulation is used in LoRa, which enables signals with different spreading factors SF $\in \mathcal{S} = \{7, ..., 12\}$ to be distinguished and received simultaneously, even if they are transmitted at the same time and on the same channel. Lower SFs lead to higher transmission rates and shorter transmission time but require a higher SNR (Signal to Noise Ratio). The sensitivity of LoRa transceivers and the reception threshold are given in Table I. Following [3], a collision occurs when two or more devices select the same channel and spreading factor. However, perfect orthogonality is not guaranteed, and interference among communications using different SF, called inter-SF collision, must be accounted for [4]. In fact, the GW can successfully receive a signal using SF s if its power is higher by a given threshold (given in Table I) than the total interference generated by concurrent signals using SF $s' \neq s$.

Furthermore, if there are several signals transmitted with the same SF and on the same channel simultaneously, the GW is still able to successfully receive the strongest signal if its SINR is higher than a threshold of 6 dB. This phenomenon is known as the capture effect [2], [3]. Therefore, apart from considering the collisions due to selecting the same SF and channel, we also consider the impact of the inter-SF collision and the capture effect for their relevance on LoRaWAN performances. Besides selecting a SF and a channel, each end-device selects a transmission power between 2 dBm and 14 dBm. Due to the space limitation of the paper, we assume that all end-devices use the same channel, and with maximal transmission power.

Table I: LoRa characteristics at BW = 125 kHz [1], [4]

SF	**Bit- rate [kbps]**	**Receiver Sensitivity [dBm] [1]**	**Reception Thresh. [dB]**	**Inter-SF collision Thresh. [dB] [4]**
7	5.47	-123	-6	-7.5
8	3.13	-126	-9	-9
9	1.76	-129	-12	-13.5
10	0.98	-132	-15	-15
11	0.54	-134.5	-17.5	-18
12	0.29	-137	-20	-22.5

We note that adopting the propagation model in [4] renders a small scale network where the signal range attains 4.5 Km. This model suits the paper purpose in obtaining a high device density akin to that of LoRaWAN while using a relatively small number of devices ($N = 100$).

To ease the performance assessment and the analysis of packet collision, we assume that all end-devices have the same packet generation rate of λ packets per hour and that all packets have the same length of l bytes. As stated in the LoRaWAN specifications [1], after sending a packet, the end-device waits for an acknowledgment (ACK) sent by the GW. We assume that there is no collision between the ACK and uplink packets. In fact, the ACK can be delivered on a separate channel with a higher duty cycle. Hence, if an end-device receives an ACK for its transmitted packet, then either there was no collision, or the capture effect has occurred. Conversely, when ACK is not received, either the packet was lost due to collision with another packet transmitted with the same SF, or due to the inter-SF collision.

III. Optimal Proportional Fair Spreading Factor Selection in LoRaWAN

In this section, the spreading factor selection is casted as an optimization problem for LoRaWAN, steered by the GW or the Network Server. A centralized solution to the SF selection problem is complex and necessitate signaling that will drain the energy of LoRa devices, supposed to have ultra-durable battery life. Thus, the centralized solution will be used as a benchmark for the distributed learning-based algorithm sketched below. Note that all end-devices have the same packet generation rate of λ packets per hour, and the same packet length of l bytes. Transmission attempts occur according to a Poisson distribution of parameter λ. We denote by N_s the maximum number of devices that can use SF s and above. Let T_s be the time needed to transmit a packet of l bytes on SF s (time on air). Then, given a duty cycle limitation of $d = 1\%$, the packet generation rate for each device operating on SF s must verify $\lambda T_s \leq d = 1\%$ [1].

We suppose that we have an external traffic (*e.g.* devices belonging to a different operator) of intensity λ_s^e packets per second on spreading factor s. Let p_s be the ratio of devices using SF s and above. We can write the normalized channel traffic on SF s as follows:

$$G_s = (\lambda \cdot N \cdot p_s + \lambda_s^e)T_s \tag{1}$$

LoRaWAN uses a simple ALOHA-based algorithm without sensing, doing away with synchronization and access reservation. Therefore, according to the Poisson traffic arrival, the normalized total throughput $\overline{G}$ of the network is given by:

$$\overline{G} = \sum_{s=1}^{S} G_s \exp(-2G_s) \tag{2}$$

We consider a network utility under proportional fairness for the normalized throughput of the network. While conventional resource allocation usually aims at maximizing the total normalized throughput in (2), it may deprive devices far away from the GW from having fair access to radio resources. Hence, in this work, we privilege the device's interest by relying on the proportional equity incarnated by the logarithmic function as in [11]. Accordingly, the spreading factor selection problem consists in computing the ratios p_s that maximize the following utility function:

$$U = \sum_{s=1}^{S} \log\left(G_s \exp(-2G_s)\right) \tag{3}$$

Such a utility function ensures a proportional fair normalized throughput, which strikes a good balance between fairness and efficiency. The optimization problem is as follows: (4):

$$(\mathcal{P}): \quad \max_{p_s} \sum_{s=1}^{S} \log\left(G_s \exp(-2G_s)\right) \tag{4a}$$

$$\text{subject to} \quad \sum_{s=1}^{S} p_s \leq 1, \tag{4b}$$

$$\sum_{i=1}^{S} p_i \leq \sum_{i=1}^{S} \frac{N_i}{N}, \forall s = 1, \ldots, S. \tag{4c}$$

The utility maximization objective is subject to constraint (4b) ensuring that the sum of ratios does not exceed 100%. Constraints (4c) ensure that the number of devices selecting SF s and above does not exceed the maximum number N_s for each SF s. The optimization problem (4) is convex with a concave objective function and linear constraints.

IV. Distributed Learning for Spreading Factor Selection in LoRaWAN

We describe the fully distributed learning-based algorithm suitable for LoRaWAN. Any end-device is considered as an intelligent agent that needs to choose at a iven time t a convenient spreading factor SF s or equivalently a strategy $s(t) = \{SFs\}$. Let $\mathcal{S} = \{7, \ldots, 12\}$ be the set of spreading factors. We consider a realistic setting where devices are unaware of their position and channel conditions, and thus unaware of their minimal SF. Therefore, they will select any SF $s \in \mathcal{S}$. Accordingly, the strategy space of any device is $\mathcal{S}$. At each iteration t (at packet arrival), each device selects a strategy $s(t)$ governed by some distribution over $\mathcal{S}$, which yields a reward $r_s(t) \in \{0, 1\}$. Successful packet transmission (acknowledged by the GW) yields $r_s(t) = 1$. In case of packet loss, $r_s(t) = 0$.

Such type of learning corresponds to the framework of the Multi-Armed Bandit (MAB) problem [8] that only makes use of local information available at the LoRaWAN end-device level (received ACK). The result of the devised algorithm in each device will be a set of SFs that suffers the least from collisions. As the distributed selection of the best radio resources by uncoordinated devices is appropriately modeled by the adversarial MAB problem, we have recourse to the popular EXP3 algorithm [9], [10]. However, the EXP3 algorithm has an exponential complexity with the size of the strategy set, leading to prohibitive convergence times. Thus, we adopt a computationally efficient version of EXP3, known as the EXP3.S [10], to determine the best SF selection.

At each iteration t (at packet arrival), each device j selects a strategy $s(t)$ with distribution $p_s^j(t)$ over $\mathcal{S}$, which renders reward $r_s(t)$. The goal of any device j is to update $p_s^j(t)$ in order to get the largest reward at horizon T in comparison with the best fixed strategy. We initialize the algorithm with all weights equal to 1, and with the uniform distribution $p_s^j(0) = \frac{1}{K}$, where K is the cardinal of strategy set $\mathcal{S}$. Further, in Algorithm 1, e is the base of the natural logarithm, i.e., $e \approx 2.7182818\ldots$, and α is an input parameter used to adjust the weights at each iteration t. Note that in case of packet loss, $r_s(t) = 0$ and no update will take place for the distribution strategy, and hence, no learning either.

V. Performance Evaluation

We consider a LoRaWAN-type network with 1 GW and $N = 100$ end-devices distributed in a disc of radius 4.5 km. When we only consider path loss, the network is composed of concentric discs corresponding to different receiver sensitivity values (given in Table I) and hence to different minimal spreading factors. Accordingly, the closest devices to the GW have a choice spanning from 7 to 12, whereas the furthest away devices are constrained with $SF = 12$, as shown in Figures 1 and 2. In our simulations, we consider also the log-distance path loss model with flat fading, where the reference distance $d_0 = 40$m, the path loss at the reference distance $PL_0 = 107.41$ dB.To evaluate the impact of devices distribution in the network, we consider two scenarios: a uniform distribution of devices, and a non-uniform distribution where we choose at random to crowd a given region (with $SF = 10$).

We will evaluate the EXP3.S performance in a real setting that accounts for both capture effect and inter-SF collision. Further, to fully assess the reinforcement learning based algorithm, we will compare EXP3.S against the fair centralized algorithm presented in Section III, but also against simple algorithms such as i) the uniform SF distribution where each device selects the SF according to a uniform distribution over $\mathcal{S}$, and ii) the random distribution where each device selects the SF according to a Gaussian distribution.

Input : Let SF $s \in \mathcal{S}$ be the strategy chosen by device j.

Initialization:

- Set the initial weights $\omega_s^j(0) = 1, \forall s \in \mathcal{S}, \forall j \in \mathcal{N}$ and the uniform distribution of strategies per device $p_s^j(0) = \frac{1}{K}$.
- Set the learning rate $\gamma = \min\left\{1, \sqrt{\frac{K \log(KT)}{T}}\right\}$.
- Set the input parameter $\alpha = 1/T$.

for $t = 1$ *to* T **do**

initialization ;

foreach *end-device* j **do**

At time t, draw strategy $s \in \mathcal{S}$ according to the distribution $p_s^j(t)$;

if *Transmit* **then**

Receive reward:

$$r_s^j(t) = \begin{cases} 1 & \text{if ACK is received,} \\ 0 & \text{otherwise.} \end{cases}$$

Update:

$$\omega_s^j(t+1) = \omega_s^j(t) \exp\left(\frac{\gamma r_s^j(t)}{K \cdot p_s^j(t)}\right) + \frac{e\alpha}{K} \sum_{s=1}^{S} \omega_s^j(t)$$

$$p_s^j(t+1) = (1-\gamma) \frac{\omega_s^j(t+1)}{\sum_{s=1}^{K} \omega_s^j(t+1)} + \frac{\gamma}{K}$$

end

end

end

Algorithm 1: EXP3.S algorithm for fully distributed SF allocation in LoRaWAN

We develop a discrete-event simulator in Python with the Simpy library [12]. It is a flexible simulation tool that captures specific LoRa link behavior for multiple network settings with the impact of capture effect and inter-SF collision. For each scenario, the time horizon for simulation is set to $T = 10^7$ iterations. The 1% LoRaWAN duty cycle limitation [1] is respected by setting the packet generation rate of each end-device to $\lambda = 15$ packet/hour. Packets are generated with exponential interarrival. The other simulation parameters are presented in Table II.

Table II: Parameters for performance analysis.

Parameters	Values
Area	Disc of radius 4.5 km
Packet length	50 bytes
Bandwidth (BW)	125 kHz
Code rate	4/5
Frequency set	868100 Hz
Capture Effect Threshold	6 dB
Inter-SF Collision Threshold	Table I
Transmission Power	14 dB

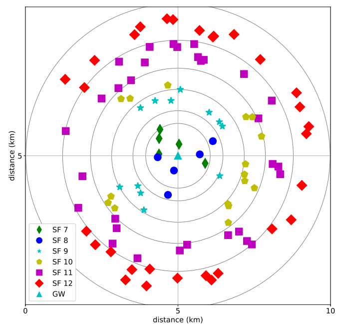


Figure 1: Impact of uniform distribution on SF selection

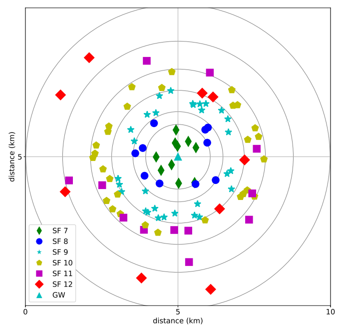


Figure 2: Impact of non-uniform distribution on SF selection

A. Spreading factor selection for EXP3.S

Figures 1 and 2 display the choice of spreading factors according to the EXP3.S algorithm 1 by devices uniformly and non-uniformly distributed respectively. The results are obtained at the end of horizon time ($T = 10^7$ iterations). One sees that the choice of spreading factors by each end-device depends on its location and on the distribution of other devices in the network (proximity of other devices).

For the uniform distribution, devices in outer regions, *i.e.*, regions which receiver sensitivity corresponds to spreading factors equal to or higher than 8, usually choose the SF corresponding to their region (their smallest feasible SF, yielding the highest bit rates). Conversely, devices in the region of SF 7 load balance their traffic between spreading factors 7 and 8, depending on their distance to the GW and the relative distances of other competing devices. In fact, by displaying in Figure 3 the strategy evolution of two randomly chosen devices, we can see that one device (shown in the upper figure) favors $SF = 7$ over $SF = 8$ (with probability 0.8), while the other device (shown in the lower figure) equally shares its traffic on both SFs.

For the non-uniform distribution, the same trend is observed in the central region with minimal $SF = 7$, but more importantly, we can see that devices in the crowded region

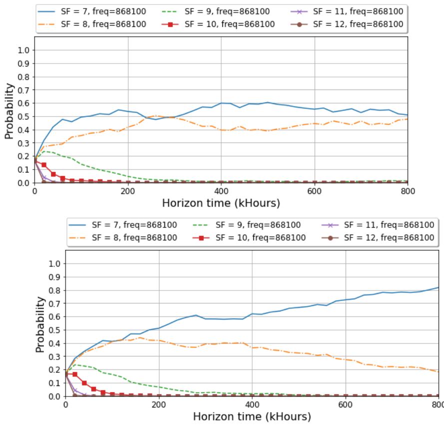


Figure 3: Uniform distribution: strategy evolution for two devices in central region

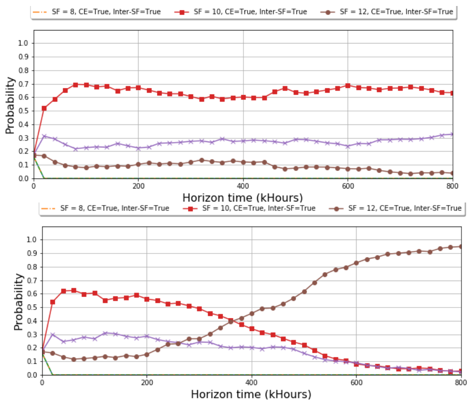
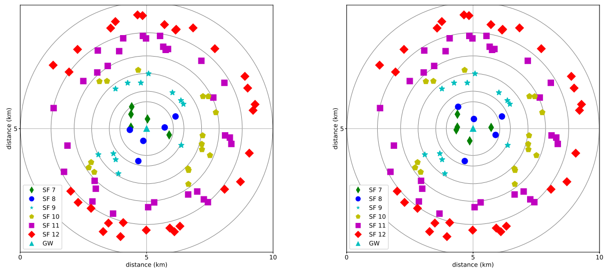


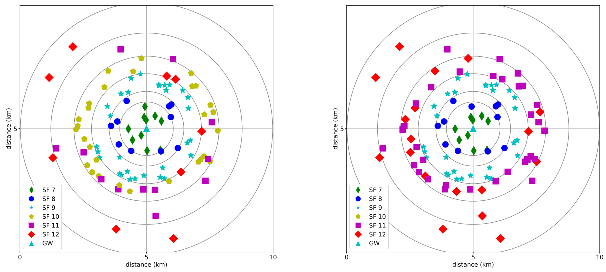
Figure 4: Non-uniform distribution: strategy evolution for two devices in crowded region

partake their traffic on their three feasible SFs: 10, 11 and 12. This behavior shows the relevance of an intelligent radio resource allocation. In particular, we display in Figure 4 the strategy evolution of two randomly chosen devices in region of $SF = 10$, that are closeby. We can see that the first device (in the upper figure) chooses $SF = 10$ with probability 0.7 and $SF = 11$ with probability 0.3, while the other device (in the lower figure), that initially took similar decisions, finally opted for $SF = 12$ to shield itself from the harmful interferences generated by its direct neighbour.

In both figures 3 and 4, we note the convergence of the EXP3.S algorithm for spreading factor selection. Convergence times are long, in the order of 30 kHours. We notice that convergence for devices in the outer region is faster than for those in the inner regions, because the latter have more feasible strategies. Finally, to gauge the impact of capture effect, we show in Figure 5 the SF constellation for the uniform distribution (upper figures 5b and 5a) and non-uniform distribution (lower figures 5d and 5c). We can see that in both geographical distributions, devices load balance their traffic on more SFs when the CE taken into account, which enhances performances



(a) Uniform distribution with CE (b) Uniform distribution w/o CE



(c) Non-uniform dist. with CE (d) Non-uniform dist. w/o CE

Figure 5: A snapshot of SF selection for EXP3.S algorithm at $t = 10^7$, w/o CE

as will be highlighted in the next subsections.

B. Successful Transmission Rate

In this subsection, we evaluate the rate of successfully received packets. In order to gain more insight on the impact of intelligent devices with learning SF capabilities on the performance of LoRaWAN, we consider three scenarios with three different ratios of intelligent devices where 0%, 50% and 100% of end-devices use EXP3.S algorithm for their spreading factor selection. Non-intelligent devices adopt either a uniform strategy or a random strategy for SF selection.

Figure 6 shows the packet reception rate PRR for the network in presence of capture effect and inter-SF collision. We can see clearly that the packet reception rate of the system with distributed learning is significantly increased compared to the uniform SF selection and random SF selection. In addition, the larger the number of intelligent end-devices using distributed learning, the higher the packet reception rate. We note that in the uniform distribution, the PRR gets close to 0.9, while in the non-uniform case, PRR surpasses it owing to the increased efficiency brought by astute SF allocation. Furthermore, we observe that taking into account CE and inter-SF collision leads to a small increase in the packet reception rate. This increase is small since the network is sparse. However, when the device density increases, the impact of CE and inter-SF collision will increase, leading to a scalability limit.

Recall that convergence times are long, in the order of 30 kHours. However, the packet reception rate PRR of the

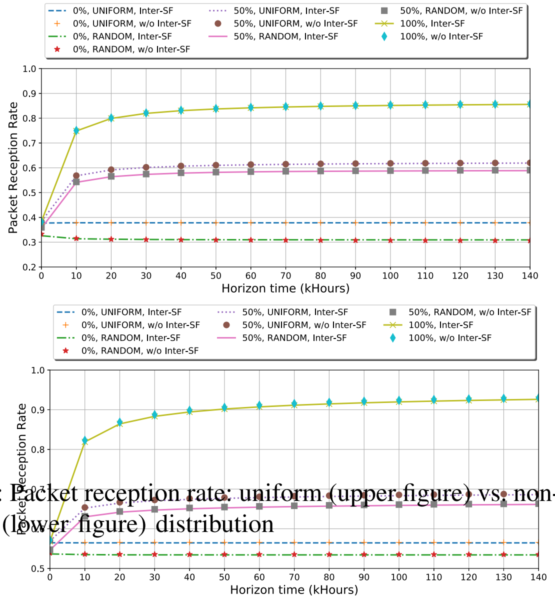


Figure 6: Packet reception rate: uniform (upper figure) vs. non-uniform (lower figure) distribution

network with distributed learning algorithm can reach 0.8 for the uniform case and 0.9 for the non-uniform case in an acceptable time (less than 10 kHours). Such time duration is acceptable for a static setting that is common in wide area IoT scenarios (in automated factories, smart cities, smart agriculture, etc.).

C. Normalized Total Throughput

In the present subsection, we evaluate the total normalized throughput, given in (2). Here, we compare the total normalized throughput with that obtained with our fair proportional optimal solution devised in (4). Figure 7 displays the average normalized total throughput of LoRAWAN, with uniform and non-uniform device distributions. We can see that the EXP3.S algorithm shows small discrepancy with the optimal solution when all devices are intelligent. Finally, we can see in the non-uniform device distribution, the performance enhancement brought by taking into account the inter-SF collision.

VI. Conclusion

In this paper, we investigated the pertinence of intelligent radio resource allocation in LoRaWAN. We put emphasis on spread spectrum allocation through reinforcement-based learning, in a realistic setting that accounts for the capture effect, collisions among spreading factors and non-uniform device distribution. In particular, we applied the EXP3.S algorithm to autonomously steer the decision of each device towards the least congested SFs while ensuring reactivity to the possible changes that can occur in the common resource usage. Further, we devised an optimal fair centralized SF allocation problem to use as a benchmark for the fully distributed EXP3.S algorithm. Extensive simulations show that the distributed

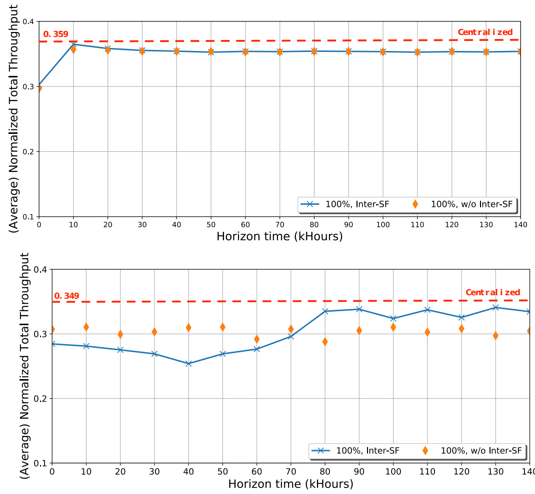


Figure 7: Normalized Throughput: uniform (upper figure) vs. non-uniform (lower figure) distribution

learning-based algorithm outperforms simple heuristics, and shows small discrepancy with the centralized optimal solution in terms of normalized total throughput.

Acknowledgment

This research was partially supported by Labex DigiCosme (project ANR-11-LABEX-0045-DIGICOSME) operated by ANR as part of the program "Investissement d'Avenir" Idex Paris-Saclay (ANR-11-IDEX-0003-02).

References

[1] O. Georgiou and U. Raza, "Low power wide area network analysis: Can lora scale?" *IEEE Wireless Comm. Letters*, vol. 6, pp. 162–165, 2017.
[2] 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, p. 150597, Oct 2015.
[3] T. Voigt, M. Bor, U. Roedig, and J. Alonso, "Mitigating inter-network interference in lora networks," in *Proceedings of the 2017 International Conference on Embedded Wireless Systems and Networks*, ser. EWSN. USA: Junction Publishing, 2017, pp. 323–328.
[4] D. Croce and et al, "Impact of spreading factor imperfect orthogonality in lora communications," *Communications in Computer and Information Science*, pp. 165–179, 2017.
[5] A. Waret, M. Kaneko, A. Guitton, and N. E. Rachkidy, "Lora throughput analysis with imperfect spreading factor orthogonality," *IEEE Wireless Comm. Letters*, pp. 1–1, 2018.
[6] R. Bonnefoi and et al, "Multi-armed bandit learning in iot networks: Learning helps even in non-stationary settings," *Cognitive Radio Oriented Wireless Networks*, pp. 173–185, 2018.
[7] A. Azari and C. Cavdar, "Self-organized low-power iot networks: A distributed learning approach," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–7.
[8] S. Bubeck and A. Slivkins, "The best of both worlds: Stochastic and adversarial bandits," in *Proceedings of the 25th Annual Conference on Learning Theory*, vol. 23. PMLR, 25–27 Jun 2012, pp. 42.1–42.23.
[9] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multi- armed bandit problem," *Machine Learning*, vol. 47, no. 2/3, pp. 235–256, 2002.
[10] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM Journal on Computing*, vol. 32, no. 1, pp. 48–77, 2002.
[11] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, Jan 1997.
[12] D.-T. Ta, K. Khawam, S. Lahoud, C. Adjih, and S. Martin, "Lora-mab: A flexible simulator for decentralized learning resource allocation in iot networks," 019 12th IFIP Wireless and Mobile Networking Conference (WMNC), 2019.