

Le retour groupé des bandits multi-bras pour aider les réseaux LoRaWAN

Jules COURJAULT¹ Baptiste VRIGNEAU¹ Olivier BERDER¹ Claude GUICHAOUA² Amy L. MURPHY³

¹Univ Rennes, IRISA, CNRS, Lannion, France

²CG Wireless, Lannion, France

³Fondation Bruno Kessler, Trento, Italy

Résumé – L’essor de l’IoT ces dernières années a conduit au déploiement massif de dispositifs connectés, rendant les réseaux IoT plus denses. Afin d’optimiser la transmission et de limiter la congestion dans ces réseaux denses, des algorithmes de gestion adaptative du débit de données ont été mis en place, comme celui utilisé dans le protocole LoRaWAN. L’utilisation d’algorithmes basés sur l’apprentissage par renforcement, en particulier les bandits manchots, a été étudiée. Cependant, la limitation du cycle d’utilisation réduit les performances de ces algorithmes dans les réseaux denses jusqu’à 15%. Cet article vise à proposer une solution pour pallier ce problème lié à la limitation du rapport cyclique, en prenant la technologie LoRa comme cas d’étude. Un effort particulier a été fait sur la consommation d’énergie, et la fonction de récompense a été modifiée afin d’économiser l’énergie tout en garantissant la qualité de service. Les performances sont évaluées grâce à notre simulateur J-LoRaNeS, basé sur le langage Julia.

Abstract – The rise of IoT in the past few years has led to the massive deployment of connected devices making IoT networks denser. To optimize transmission arising from congestion in dense networks, adaptive data-rate algorithms have been implemented such as the one used in the LoRaWAN protocol. Utilization of algorithms based on reinforcement learning, especially multi-armed bandit, have been investigated, but the duty-cycle limitation decreases the performance of these algorithms in dense networks up to 15%. This paper aims at giving a solution to resolve the issue caused by duty-cycle limitation, using the LoRa technology as a study case. An effort was done on energy consumption and the reward is modified in order to save energy according to the quality of service. Performances are evaluated thanks to our simulator J-LoRaNeS based on the Julia language.

1 Introduction

L’essor de l’Internet des Objets (IoT) ces dernières années a conduit au déploiement massif de dispositifs connectés. La technologie Long Range (LoRa) [9], soutenue par Semtech et l’alliance LoRa, offre un très bon compromis entre portée, débit de données et consommation énergétique. Elle permet un débit allant de 0,3 kbps à 50 kbps dans les bandes libres 433 MHz, 868 MHz ou 915 MHz du spectre industriel, scientifique et médical (ISM), en utilisant le principe de la modulation chirp spread spectrum, offrant ainsi une transmission quasi-orthogonale.

Dans la norme LoRaWAN, un algorithme appelé Adaptive Data-Rate (ADR) est proposé afin d’adapter les paramètres de communication au canal, à savoir le facteur d’étalement (Spreading Factor, SF) et la puissance d’émission (Transmit Power, TP). Les performances de cet algorithme ont été étudiées dans [8], mettant en évidence plusieurs inconvénients, notamment un temps de convergence élevé, entraînant une augmentation des pertes de paquets et de la consommation énergétique.

Pour pallier ces limites, des algorithmes ADR basés sur les bandits manchots multi-bras (MAB), une technique d’apprentissage par renforcement (RL), ont été proposés. Ces approches permettent d’améliorer le taux de livraison des paquets (PDR), de réduire la consommation énergétique des nœuds du réseau et de diminuer le temps de convergence. Cependant, pour la bande de 868 MHz en Europe, les réglementations de l’ETSI imposent une limitation du rapport cyclique (DC) à 1% ou

10% selon la fréquence sélectionnée. Cette limitation pose problème pour l’ADR basé sur MAB, car elle réduit la rétroaction envoyée aux terminaux utilisateurs (EDs), ce qui peut entraîner des calculs de récompense incorrects et, par conséquent, de mauvaises décisions d’action pour les paramètres de communication sélectionnés.

Il est cependant possible de tirer parti des techniques MAB tout en atténuant l’effet négatif de la contrainte de rapport cyclique. La solution proposée dans cet article consiste à regrouper les rétroactions effectuées par la passerelle (GW) vers les EDs. Cela réduit l’utilisation du DC par la GW et permet à l’agent d’exploiter une récompense fiable.

2 Adaptation de débit à l’aide du MAB

2.1 Principes de base du MAB

Les algorithmes MAB font partie du paradigme d’apprentissage par renforcement en intelligence artificielle. Ainsi, ils interagissent avec leur environnement afin d’apprendre la meilleure décision (bras) à prendre. Lors de la réception d’une récompense $r[t_k]$, la politique π est mise à jour, puis un bras $a[t_k]$ est choisi en fonction de cette politique mise à jour, ce qui conduit à une nouvelle récompense $r[t_{k+1}]$.

Lorsqu’un algorithme MAB est utilisé pour remplacer l’algorithme ADR de LoRaWAN, la récompense est généralement composée d’une composante énergétique combinée à une mesure de la qualité de service telle que le PDR. Dans la suite de

cet article, la récompense est calculée comme suit :

$$r = \begin{cases} \alpha \frac{EC_{\max} - EC}{EC_{\max} - EC_{\min}} + 1 - \alpha, & \text{si ACK reçu} \\ 0, & \text{sinon} \end{cases}, \quad (1)$$

où EC est l'énergie consommée par le paquet, EC_{\min} , EC_{\max} correspondent respectivement au consommation minimale et maximale parmi toutes les configurations possibles, et $\alpha \in [0; 1]$ est un poids énergétique utilisé pour donner plus ou moins d'importance à la consommation d'énergie [5].

2.2 Travaux existants

Les auteurs de [6] ont étudié les performances d'un seul ED utilisant un algorithme MAB dans un réseau où les autres EDs emploient l'algorithme ADR standard de LoRaWAN. Les actions disponibles (ou bras du bandit) correspondent aux mêmes paires (SF, TP) que celles utilisables par l'ADR de LoRaWAN. Les auteurs ont testé ces configurations avec l'algorithme Exponential-weight for Exploration and Exploitation (EXP3) et ses variantes [1], les bornes de confiance supérieures (Upper Confidence Bounds, UCB) [3] ainsi que différentes variantes de l'algorithme Thompson Sampling (TS) [10]. Leurs simulations montrent une réduction drastique du nombre de paquets perdus et de la consommation énergétique pour tous les algorithmes MAB testés.

Une version modifiée de l'algorithme EXP3 est étudiée dans [11]. Bien que cette modification permette d'éviter l'utilisation de bras ayant de faibles performances, elle représente une solution risquée dans des environnements susceptibles de changer, car l'algorithme pourrait ne pas réussir à trouver une solution après une variation du canal.

Les auteurs de [12] ont mené des expériences en utilisant la dynamique Tug of War [7], l'algorithme Upper Confidence Bound 1 [4] et l'algorithme ϵ -Greedy pour sélectionner une valeur de facteur d'étalement (SF) ainsi qu'un canal dans la bande ISM de 920 MHz au Japon, dans un réseau composé d'une passerelle (GW) et jusqu'à 30 noeuds (EDs). Les résultats montrent que la sélection conjointe d'un canal et d'une valeur de SF offre de meilleures performances que l'approche indépendante.

Les auteurs de [2] présentent enfin un algorithme basé sur l'algorithme Discounted UCB, capable de détecter les changements dans les propriétés stochastiques des récompenses.

3 Retour d'Information Groupé

3.1 Problématique liée au rapport cyclique

La Fig. 1 montre l'évolution du PDR pour différents algorithmes ADR, en considérant différentes hypothèses pour le MAB : *Oracle* et *DC-constrained*. Dans le cas *Oracle*, il est supposé qu'un paquet DL est envoyé après la réception d'un paquet UL, tandis que dans le cas *DC-constrained*, le paquet DL n'est envoyé que si la limitation DC peut être respectée. Le PDR est calculé sur une fenêtre glissante d'une heure avec un chevauchement de 50%.

Nous pouvons observer que les algorithmes MAB ont des temps de convergence beaucoup plus courts, et que le PDR obtenu après convergence est supérieur à celui de l'ADR de LoRaWAN dans les versions *Oracle* et *DC-constrained*. D'autre part, comme prévu, dans le cas *DC-constrained*, les algorithmes

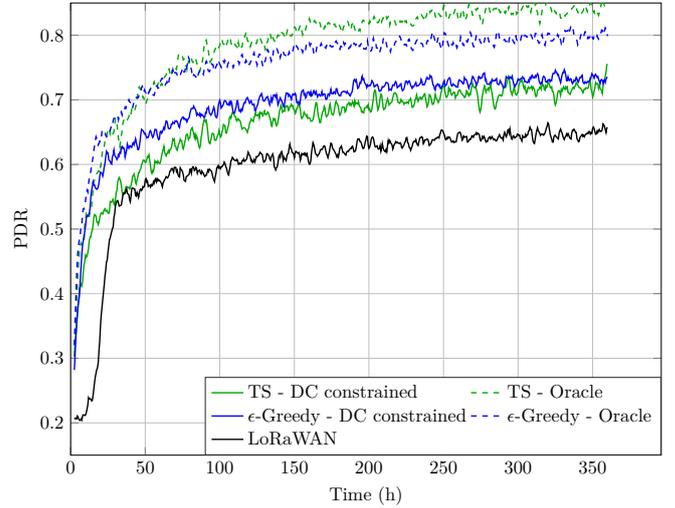


FIGURE 1 : Taux de livraison des paquets pour les algorithmes ϵ -Greedy et TS, dans les cas DC-constrained et Oracle, comparé à LoRaWAN ADR.

MAB sont moins performants que dans le cas *Oracle* (80% pour ϵ -Greedy et 85% pour TS à la fin de la simulation pour les cas *Oracle* contre 72% pour les deux algorithmes en *DC-constrained*). Le PDR de l'ADR LoRaWAN, quant à lui, ne dépasse jamais 65%.

Le respect de la contrainte de *Duty Cycle* (DC) au niveau de la GW entraîne donc une perte de performance pour l'algorithme MAB, en raison d'un calcul de récompense qui ne correspond pas à l'état de réception du paquet UL. Cela est illustré dans la Fig. 2, qui montre les différents cas pouvant survenir lorsqu'un paquet UL est envoyé par un ED.

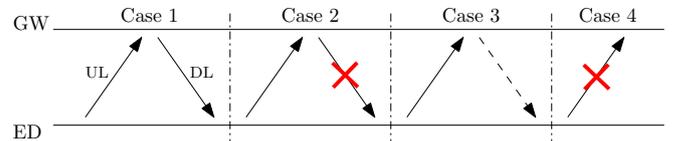


FIGURE 2 : Échange de paquets entre un ED et une GW. Lorsqu'un ED envoie un UL, quatre cas peuvent être observés. Cas 1 : L'UL est reçu, un DL est envoyé par la GW et correctement reçu. Cas 2 : L'UL est reçu, un DL est envoyé, mais il n'est pas reçu par l'ED. Cas 3 : L'UL est reçu, mais aucun DL n'est envoyé en raison de la contrainte de DC. Cas 4 : L'UL n'est pas reçu par la GW.

3.2 Contourner la limitation du DC en groupant le retour d'informations

Pour limiter l'activité de la passerelle et respecter la contrainte de transmission en DL, nous proposons de regrouper les retours d'information adressés aux ED. La Fig. 3 illustre l'algorithme proposé : l'agent sélectionne un bras selon la politique choisie¹ pour chaque UL. L'ED a une probabilité $p \in [0, 1]$ d'effectuer une demande de DL lors de l'envoi d'un UL, c'est-à-dire, de positionner le bit ADR ACK^2 à 1. Lorsqu'un UL

¹ ϵ -Greedy, TS, EXP3, etc. Tout algorithme MAB classique peut être utilisé.

²Ce bit est déjà disponible dans le standard LoRaWAN.

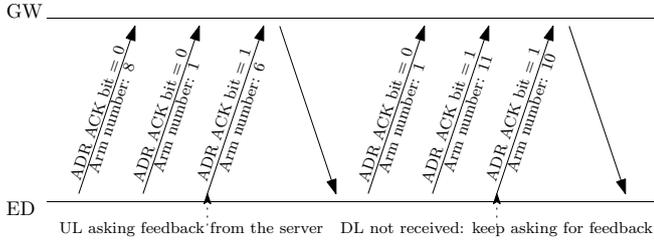


FIGURE 3 : Exemple illustratif de l’algorithme MAB avec retour d’information groupé.

avec une demande de retour est reçu, un DL est envoyé avec un retour d’information indiquant quels paquets UL ont été reçus depuis le dernier envoi d’un DL. À la réception du DL, l’agent met à jour sa politique en fonction des informations reçues.

En cas de non réception du DL, une nouvelle demande de retour d’information est effectuée, jusqu’à réception du DL. Cependant, selon la raison de la non-réception du DL, les conséquences sur l’algorithme varient. Si l’UL n’est pas reçu, ou si la GW ne peut pas envoyer de DL en raison de la contrainte de DC, aucun impact majeur ne sera observé. En revanche, si le DL est envoyé mais non reçu, l’agent ne pourra pas mettre à jour sa politique, entraînant un retard dans l’apprentissage.

4 Résultats de simulation

Nos simulations sont réalisées à l’aide du simulateur J-LoRaNeS [5], qui permet l’utilisation des algorithmes MAB ainsi que le suivi de diverses métriques telles que le taux de livraison des paquets (PDR), la consommation d’énergie moyenne et le *duty cycle* au niveau de la passerelle.

En raison de leur faible complexité, les algorithmes ϵ -greedy et TS sont considérés pour l’ADR. Les configurations possibles sont les mêmes que celles de l’ADR LoRaWAN, *i.e.*, lorsque $SF = 7$, toutes les valeurs de puissance de transmission (TP) peuvent être utilisées, tandis que pour $SF \geq 8$, seule la puissance maximale peut être utilisée.

Le réseau d’évaluation comprend une seule GW et 500 ED de classe A, positionnés aléatoirement dans une zone carrée de $10 \text{ km} \times 10 \text{ km}$. Chaque ED émet un seul paquet de 20 octets à un moment aléatoire dans chaque intervalle de 10 minutes. La configuration de fréquence choisie permet un DC de 1% pour les UL et les DL envoyés sur RX1, tandis que la limite de DC pour les DL envoyés sur RX2 est de 10%³. Le modèle de canal d’Okumura-Hata pour une petite ou moyenne ville est utilisé pour les simulations. Dans les sections suivantes, le terme "classique" fait référence aux algorithmes MAB qui n’utilisent pas le retour d’information groupé (GF) décrit précédemment.

4.1 Analyse du rapport cyclique

La Table 1 présente le rapport cyclique de la station de base pour différentes tailles de réseau (N_{ED} représente le nombre de noeuds) en utilisant l’algorithme ϵ -Greedy avec un retour

³RX1 et RX2 sont les fenêtres d’écoute des ED autorisées par le standard LoRaWAN pour les ED de classe A

TABLE 1 : PDR et DC obtenus en utilisant l’algorithme MAB ϵ -Greedy classique ou avec retour groupé, pour différents N_{ED} .

| N_{ED} | | 50 | 250 | 350 | 500 |
|-----------|------------|-------|-------|-------|-------|
| Classique | PDR | 0.980 | 0.906 | 0.870 | 0.789 |
| | DC_{RX1} | 0.472 | 0.882 | 0.925 | 0.954 |
| | DC_{RX2} | 2.711 | 7.892 | 8.398 | 8.840 |
| Groupé | PDR | 0.984 | 0.920 | 0.893 | 0.842 |
| | DC_{RX1} | 0.353 | 0.811 | 0.866 | 0.908 |
| | DC_{RX2} | 1.732 | 6.901 | 7.715 | 8.371 |

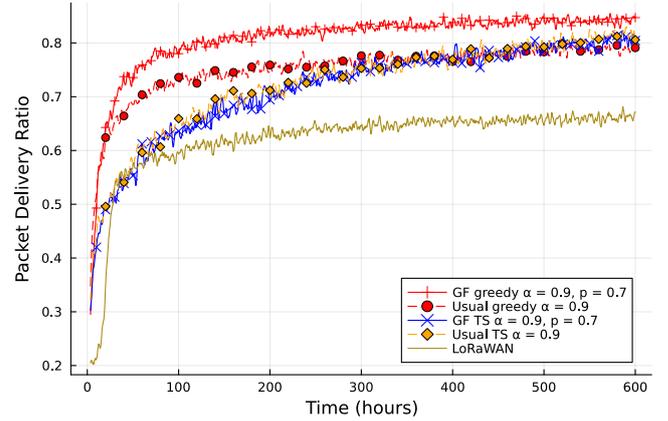


FIGURE 4 : PDR moyen pour les algorithmes ϵ -Greedy et TS classique et utilisant le retour groupé avec le couple (α, p) après une recherche numérique.

d’information classique ou groupé. On remarque que l’utilisation du retour groupé permet effectivement de réduire sensiblement le rapport cyclique.

4.2 Taux de paquets reçus

Une recherche exhaustive a montré que le couple $(p = 0.7, \alpha = 0.9)$ maximise le PDR du réseau dans notre configuration. Lors de cette recherche, nous avons constaté que l’augmentation de α tend à une augmentation du PDR du réseau. Ce résultat contre-intuitif s’explique par le fait qu’une valeur plus élevée de α favorise les petites valeurs de SF réduisant ainsi les risques de collisions de paquets. La Fig. 4 présente le PDR moyenné sur une fenêtre glissante d’une heure avec un chevauchement de 50%, pour les algorithmes MAB ϵ -Greedy et TS, en configurations standard et GF, avec $\alpha = 0.9$, correspondant à la configuration offrant les meilleures performances pour les deux algorithmes MAB.

On note que chaque configuration atteint un PDR plus élevé que l’ADR de LoRaWAN à la fin de la simulation. Le PDR est de 60% pour LoRaWAN, d’environ 80% pour l’algorithme ϵ -Greedy classique ainsi que les deux variantes de TS, et de 85% pour l’algorithme ϵ -Greedy utilisant GF, ce qui correspond approximativement au PDR du cas oracle présenté dans la Fig. 1.

De plus, on peut observer que le temps de convergence de l’algorithme TS est plus long que celui de ϵ -greedy. En effet, à la fin de la simulation, ϵ -greedy a convergé et devrait atteindre 90% de sa valeur finale après 75h, tandis que l’algorithme TS est encore en phase d’apprentissage à la fin de la simulation

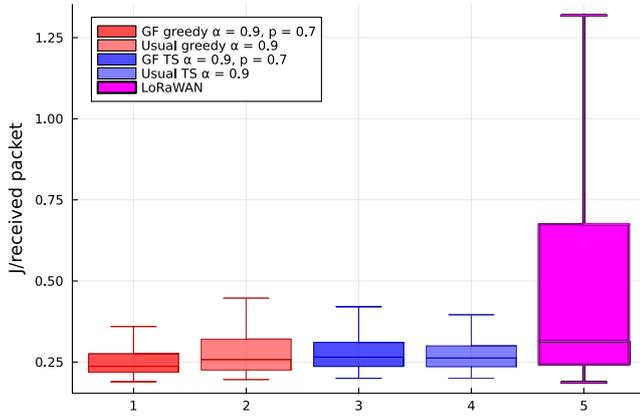


FIGURE 5 : Énergie consommée moyenne pour les différents algorithmes testés.

et nécessite 230h pour atteindre 90% de la valeur du PDR obtenue à la fin de la simulation.

4.3 Compromis Énergie - Qualité de service

La Fig. 5 présente un diagramme en boîte basé sur la consommation énergétique moyenne par paquet UL réussi pour chaque ED du réseau. On peut observer que l'utilisation de l'algorithme MAB permet une diminution importante de la consommation d'énergie par rapport à l'ADR de LoRaWAN. En effet, la consommation d'énergie médiane pour l'ADR basé sur MAB est d'environ 0,25 J par paquet UL reçu, tandis qu'elle est de 0,31 J pour l'ADR de LoRaWAN avec un écart-type élevé.

De plus, la technique GF permet de réduire la médiane et l'écart-type de la consommation d'énergie moyenne par rapport à l'utilisation classique du MAB pour l'algorithme ε -greedy. En revanche, le retour groupé n'apporte pas d'amélioration à l'algorithme TS.

5 Conclusions et perspectives

Dans cet article, une solution à la contrainte de rapport cyclique (DC) au niveau de la passerelle est proposée en regroupant les accusés de réception envoyés aux nœuds du réseau à l'aide d'un algorithme ADR basé sur les MAB et simulé grâce à J-LoRaNES. Cette solution simple permet de calculer la récompense en utilisant l'état de réception du paquet UL, rendant ainsi le calcul de la récompense plus fiable pour la mise à jour de la politique. De plus, elle réduit le DC au niveau de la GW, rendant celle-ci plus disponible lorsqu'un paquet DL doit être envoyé. Cette solution peut être mise en œuvre sans ajouter une grande complexité au système, car elle repose sur des mécanismes déjà présents dans l'ADR de LoRaWAN.

Les performances de cette solution peuvent être étudiées pour des réseaux hétérogènes. En effet, les simulations classiques considèrent des nœuds identiques, mais qu'en est-il du comportement avec une variété de nœuds (différentes périodes de transmission, différents mécanismes ADR, etc.) ? Par ailleurs, la réduction de l'utilisation du DC pourrait être suffisante pour permettre des retransmissions dans les applications sensibles, garantissant ainsi la réception des données.

Références

- [1] Robin ALLESIARDO et Raphaël FÉRAUD : Exp3 with drift detection for the switching bandit problem. *In IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–7, 2015.
- [2] Anjali R ASKHEDKAR et Bharat S CHAUDHARI : Multi-armed bandit algorithm policy for LoRa network performance enhancement. *Journal of Sensor and Actuator Networks*, 12(3):38, 2023.
- [3] Peter AUER, Nicolo CESA-BIANCHI et Paul FISCHER : Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47:235–256, 2002.
- [4] Amin AZARI et Cicek CAVDAR : Self-organized low-power IoT networks : A distributed learning approach. *In IEEE Global Communications Conference (GLOBECOM)*, pages 1–7, 2018.
- [5] Jules COURJAULT, Baptiste VRIGNEAU, Olivier BERDER, Yvon LEGOFF et Claude GUICHAOUA : Catching the LoRa ADR bandit with a new sheriff : J-LoRaNeS. *In Proceedings of the Int'l ACM Symposium on Mobility Management and Wireless Access*, pages 75–82, 2023.
- [6] Raouf KERKOUICHE, Réda ALAMI, Raphaël FÉRAUD, Nadège VARSIER et Patrick MAILLÉ : Node-based optimization of LoRa transmissions with multi-armed bandit algorithms. *In 25th IEEE International Conference on Telecommunications (ICT)*, pages 521–526, 2018.
- [7] Song-Ju KIM, Masashi AONO et Masahiko HARA : Tug-of-war model for multi-armed bandit problem. *In Unconventional Computation : 9th International Conference, Tokyo, Japan, June 21-25. Proceedings 9*, pages 69–80. Springer, 2010.
- [8] Shengyang LI, Usman RAZA et Aftab KHAN : How agile is the adaptive data rate mechanism of LoRaWAN ? *In IEEE Global Communications Conference (GLOBECOM)*, pages 206–212, 2018.
- [9] O.B.A. SELLER et N. SORNIN : Low power long range transmitter, Aug. 2014. US Patent App. 14/170,170.
- [10] Jonathan SHAPIRO, Carlos M CARVALHO et Pradeep RAVIKUMAR : Thompson sampling in switching environments with bayesian online change point detection. *In Artificial Intelligence and Statistics (AISTATS)*, pages 442–450. Microtome Publishing, 2013.
- [11] B. TEYMURI, R. SERATI, N.A. ANAGNOSTOPOULOS et M. RASTI : LP-MAB : Improving the Energy Efficiency of LoRaWAN Using a Reinforcement-Learning-Based Adaptive Configuration Algorithm. *Sensors*, 2023.
- [12] Ikumi URABE, Aohan LI, Minoru FUJISAWA, Song-Ju KIM et Mikio HASEGAWA : Combinatorial mab-based joint channel and spreading factor selection for lora devices. *Sensors*, 23(15):6687, 2023.