

Université Assane Seck de Ziguinchor

UFR Sciences et Technologies

Département Informatique



Mémoire de fin d'études

Pour l'obtention du diplôme de Master

Mention : Informatique

Spécialité : Réseaux & Systèmes

Sujet :

Proposition d'une solution basée sur k-means pour minimiser les interférences dans un réseau LoRaWAN.

Présenté et soutenu par :

M. Daouda NDIAYE

Sous la direction de :

Pr Ousmane DIALLO

Dr El hadji Malick NDOYE

Membres du jury

Pr Youssou DIENG	Professeur assimilé (MC CAMES)	Président	UASZ
Pr Ousmane DIALLO	Professeur assimilé (MC CAMES)	Encadrant	UASZ
Dr El hadji Malick NDOYE	Maître de conférences titulaire (MA CAMES)	Co-Encadrant	UASZ
Dr Thierno Amadou DIALLO	Maître de conférences titulaire (MA CAMES)	Examineur	UASZ
M Malaw NDIAYE	Enseignant Chercheur	Rapporteur	UASZ

Année universitaire 2022/2023

REMERCIEMENTS

Nous tenons, en premier lieu, à remercier le bon **Dieu** de nous avoir donné la chance d'exister, le courage et la patience pour réaliser ce projet.

Nous remercions nos chers parents et la famille **NDIAYE** pour leurs grands soutiens.

Nous remercions nos encadrants Monsieur **El hadji Malick NDOYE** et Monsieur **Ousmane DIALLO** pour avoir encadré et donné la chance d'acquérir des connaissances dans le domaine des réseaux de capteurs sans fil, pour leur présence, leur motivation et son sens d'organisation.

Nous remercions également **les membres du jury** pour avoir accepté de juger le travail.

Je tiens à exprimer ma profonde gratitude envers les familles **DARY, TENDENG, BADJI** et **MBALO** qui ont joué un rôle exceptionnel en tant que tuteurs. Votre soutien indéfectible, vos conseils éclairés et votre présence bienveillante ont constitué les piliers essentiels de la réussite de ce parcours académique. Merci du fond du cœur pour vos soutiens inestimables qui ont enrichi mon expérience et m'ont permis d'atteindre ce moment significatif.

Nos remerciements à mes amis, mes camarades de promos et mes camarades de chambres qui ont contribué de près ou de loin à la réalisation de ce travail.

DEDICACE

A mon défunt père Babacar NDIAYE, Ma mère Astou Kane THIAM,

Mes frères et sœurs,

Mes neveux et nièces,

Mes oncles et tantes.

RESUME

Le déploiement de l'Internet des Objets (IoT) dans des environnements étendus exige des solutions de connectivité à longue portée. Ce mémoire explore les technologies LPWAN (Low-Power Wide-Area Network) avec un accent particulier sur LoRaWAN combiné avec l'algorithme de k-means pour le clustering des dispositifs finaux. La proposition vise à appliquer l'algorithme de clustering k-means dans le contexte des réseaux LoRa (Long Range) en introduisant un mécanisme novateur d'activation de zones. L'objectif est de minimiser les interférences co-SF, d'optimiser la consommation énergétique et de permettre une activation sélective de zones en fonction des conditions environnementales. Cette approche permet une allocation plus efficace des ressources, une réponse rapide aux changements environnementaux et une minimisation des interférences.

Mots clés : IoT, LoRaWAN, k-means, activation de zones et interférences co-SF.

ABSTRACT

The deployment of the Internet of Things (IoT) in expansive environments demands long-range connectivity solutions. This thesis explores Low-Power Wide-Area Network (LPWAN) technologies with a specific focus on LoRaWAN combined with the k-means algorithm for end device clustering. The proposal aims to apply the k-means clustering algorithm in the context of LoRa (Long Range) networks by introducing an innovative zone activation mechanism. The goal is to minimize co-SF interferences, optimize energy consumption, and enable selective zone activation based on environmental conditions. This approach allows for more efficient resource allocation, swift response to environmental changes, and interference minimization.

Keywords: IoT, LoRaWAN, k-means, zone activation and co-SF interferences.

Table des matières

REMERCIEMENTS	i
DEDICACE	ii
RESUME	iii
ABSTRACT	iv
Liste des figures	viii
Liste des tableaux	ix
Liste des abréviations:	x
Introduction générale	1
Chapitre 1 : Les technologies LPWAN dans l'IoT	3
Introduction	3
I. Concept générale de l'IoT	3
A. Architecture de réseau IoT	3
B. Les technologies de communication de l'IoT	4
1. Technologie de courte portée	4
2. Technologie de longue portée haut débit	5
3. Technologie de longue portée bas débit	5
II. Les Technologies LPWAN	6
A. Les caractéristiques des technologies LPWAN	7
1. La portée	7
2. La consommation d'énergie	8
3. Les couts	9
4. L'évolutivité	9
B. Sigfox	9
C. Narrow Band-Internet of Things (NB-IoT)	10
D. LoRaWAN	10
E. Comparaison des technologies LPWAN	11
III. La technologie LoRaWAN	12
A. Architecture LoRaWAN	13
B. Les dispositifs finaux	14
C. Les interférences dans LoRa	15
D. La qualité de services (QoS) dans LoRa	16
E. La technologie LoRaWAN dans le secteur professionnel	16
1. Bâtiment intelligent	16
2. Agriculture	17
3. Ville intelligente	17
F. Aperçu de la technologie LoRa	17
1. Couche physique LoRa	17
2. Couche MAC LoRaWAN	18

Conclusion.....	19
Chapitre 2 : Les techniques de clustering dans les réseaux IoT	21
Introduction	21
I. Apprentissage automatique.....	21
II. Les types d'apprentissage automatique	22
A. Apprentissage supervisé	22
1. Classification.....	23
2. Régressions.....	23
B. Apprentissage non supervisé	23
1. Clustering.....	23
2. La réduction de la dimensionnalité.....	23
C. Apprentissage par renforcement	24
D. Comparaison entre l'apprentissage supervisé et non supervisé	24
III. Le clustering.....	26
A. Caractéristique du clustering	26
1. Propriétés des clusters	26
2. Propriétés du CH.....	27
B. Propriétés du processus de clustering	27
C. Les méthodes de clustering.....	28
IV. Les algorithmes hiérarchiques	29
A. LEACH.....	29
1. La phase d'initialisation.....	30
2. La phase de transmission.....	31
3. Les avantages et limites de LEACH	31
4. Architecture de communication de LEACH	31
B. Les variantes du protocoles LEACH.....	32
1. Avantages et limites des protocoles dérivés de LEACH	33
V. Les algorithmes classiques et flous	34
A. L'algorithme k-means.....	34
1. La sélection de cluster head (CH)	35
2. Les principes de l'algorithme k-means.....	36
3. L'algorithme de k-means.....	37
4. Les avantages et limites de k-means	38
B. L'algorithme de clustering Forel	39
C. L'algorithme de FCM	40
1. La méthode Fuzzy C_Means	40
2. Les étapes de l'algorithme FCM.....	42
3. Les avantages et les limites de Fuzzy C_Means.....	42
4. Organigramme de FCM	43
D. Les variantes de Fuzzy C_Means.....	44
1. L'algorithme de Tsai et Lin FCM- σ	44
2. L'algorithme de Kemel fuzzy C_Means KFCM.....	45

3. L'algorithme de C moyennes Possibiliste (PCM).....	46
VI. Comparaison des propriétés des algorithmes	47
A. Intra-clustering.....	47
B. Inter-clustering	48
C. Equilibrage des charges	48
D. Considération énergétique.....	49
Conclusion.....	50
Chapitre 3 : L'algorithme de clustering k-means Appliqué dans LoRa.....	52
Introduction	52
I. Utilisation de k-means dans LoRa	52
II. Les limites des solutions existantes	54
III. Problématiques	56
A. Algorithme de l'article de référence	56
B. Synthèse.....	59
Conclusion.....	60
Chapitre 4 : Proposition d'une solution basée sur k-means pour minimiser les interférences dans un réseau LoRaWAN.....	61
Introduction	61
I. Contexte et motivation	61
II. Modèle de déploiement et hypothèse du réseau.....	62
III. Etude détaillé de la proposition.....	62
A. La Notion de zone	63
B. Découpage en zone	64
C. Activation de zone	65
IV. Discussion des algorithmes proposés	69
A. L'algorithme de découpage en zone	69
B. L'algorithme de liste des zones à activer.....	71
C. L'algorithme d'activation de zone	71
Conclusion.....	73
Conclusion générale et perspectives	75
Bibliographie.....	76

Liste des figures

Figure 1: Architecture du réseau IoT [5]	4
Figure 2: Topologies en étoile, maillée et en arbre [24]	8
Figure 3: Architecture de Sigfox [21]	10
Figure 4: Architecture de NB-IoT [20]	10
Figure 5: Architecture du réseau LoRaWAN	14
Figure 6 : Capteur de Qualité des sols LoRaWAN	14
Figure 7: Pile de LoRaWAN	19
Figure 8: Les applications les plus importants de l'apprentissage automatique [34]	22
Figure 9: Les types d'apprentissage automatique [34]	22
Figure 10: Apprentissage par renforcement [34]	24
Figure 11: Les caractéristiques des algorithmes de clustering [42]	28
Figure 12: Routage hiérarchique basé sur le clustering [45]	30
Figure 13: Opération de l'étape d'initialisation de LEACH [47]	31
Figure 14: : Architecture de communication du protocole LEACH [45]	32
Figure 15: Processus de travail de LEACH amélioré [47]	32
Figure 16: Architecture TV-LEACH [49]	33
Figure 17: Formation d'un cluster	36
Figure 18: Commande des nœuds avec le numéro d'identification	36
Figure 19: Organigramme de l'algorithme de k-means [52]	37
Figure 20: Déplacement de la sphère dans la direction du point extrême [20]	40
Figure 21: Organigramme de l'algorithme FCM [53]	43
Figure 22: Distribution uniforme de N=500 dans une zone de rayon 3 km	62
Figure 23: Représentation de zone, de bande et de l'angle θ	63
Figure 24: L'algorithme de découpage en zone	64
Figure 25: L'algorithme des listes de zones à activer	66
Figure 26: L'algorithme d'activation des listes de zone	68
Figure 27: Représentation des bandes, des zones et l'angle θ	70
Figure 28: Représentation de deux zones activées	72

Liste des tableaux

Tableau 1 : Comparaison des technologies de communication de l'IoT	6
Tableau 2 : Tableau récapitulatif des technologies LPWAN	12
Tableau 3 : Vitesse de transmission et sensibilité du récepteur à partir du facteur d'étalement (SF)[2,11]	18
Tableau 4 : comparaison d'apprentissage supervisé et non [35]	25
Tableau 5 : Les algorithmes dérivés de LEACH	34
Tableau 6 : Comparaison des algorithmes de clustering.....	50
Tableau 7 : Utilisation de la série Fibonacci pour une valeur de k, le nombre de nœuds et la moyenne	58
Tableau 8 : Probabilité de succès pour chaque SF	58
Tableau 9 : le nombre de nœuds moyenne dans chaque cluster pour les différentes séries utilisées	59
Tableau 10 : Tableau de l'ensemble des zones Z.....	65
Tableau 11 : Représentation de la table de hachage	67
Tableau 12 : Tableau de l'ensemble L des sous listes li.....	67
Tableau 13 : Tableau de l'ensemble des sous listes à activer	69

Liste des abréviations

BW	Bandwidth
CAPEX	Capital Expenditure
CH	Cluster Head
CM	Cluster Member
CR	Code Rate
CSS	Chirp Spread Spectrum
DSSS	Direct Sequence Spread Spectrum
ED	End Devices GW Gateway
IoT	Internet of Things
K	Represents the number of clusters LoRa Long Range
LPWAN	Low-Power Wide-Area Network
LTE	Long-Term Evolution
LTE-M	Long-Term Evolution for Machines
NB	Narrow Band
NB-IoT	Narrow Band-Internet of Things
NS	Network Server
QoS	Quality of Service
RA	Radio Access Network
RF	Radio Frequency
RPMA	Random Phase Multiple Access
RSSI	Received Signal Strength Indicator
SB	Base Stations
SNR	Signal-to-Noise Ratio
SF	Spreading Factor
TDMA	Time Division Multiple Access
TTN	The Things Network
OPEX	Operating Expenditure
PAM	Partitioning Around Medoids

Introduction générale

Les récentes avancées technologiques ont ouvert de nouvelles possibilités dans le domaine des réseaux de capteurs sans fil. L'une de ces avancées est la technologie LoRaWAN (Long Range Wide-Area Network), qui permet la transmission longue portée de données avec une consommation d'énergie réduite. Cette technologie a suscité un intérêt croissant dans de nombreux domaines tels que la gestion urbaine, l'agriculture intelligente, la logistique et bien d'autres [1].

Dans le cadre du regroupement des dispositifs finaux similaires dans les réseaux LoRaWAN, les techniques de clustering jouent un rôle essentiel. Le clustering est une catégorie d'apprentissage automatique non supervisée qui regroupe les dispositifs finaux en différents clusters ou groupes homogènes en fonction de leurs similarités. L'algorithme k-means est populaire et largement utilisé pour la création de clusters. Il est l'une des techniques les plus utilisées qui vise à segmenter en k clusters (groupes) une collection de n objets. Il itère pour attribuer chaque donnée à l'un des k clusters prédéfinis [2].

La combinaison de l'algorithme k-means et de la technologie LoRa (Long Range) peut offrir plusieurs avantages potentiels, notamment en termes, de choix de la valeur k (cluster), de la surface de la zone, d'efficacité énergétique et de capacité à traiter des réseaux étendus.

Dans cette proposition, nous explorons l'intégration de l'algorithme de clustering k-means avec la technologie LPWAN LoRa pour minimiser les interférences co-SF (Spreading factor (SF), facteur d'étalement en français). L'interférence co-SF se produit lorsque des dispositifs utilisent le même facteur d'étalement (SF) pour transmettre des données dans un même voisinage. Chaque facteur d'étalement définit une plage de fréquences plus large ou plus étroite sur laquelle les données sont transmises, et les transmissions de différents dispositifs avec le même facteur d'étalement peuvent se chevaucher. Le facteur d'étalement est un paramètre spécifique aux dispositifs finaux LoRa, qui est utilisé pour contrôler la bande passante et la robustesse de la transmission des données sur un réseau LoRa.

L'objectif consiste à réduire les interférences co-SF tout en rationalisant la consommation d'énergie. Ceci est accompli en activant de manière sélective des zones spécifiques du réseau en réponse aux changements dynamiques des conditions environnementales ou des exigences de collecte de données. Cette approche vise à optimiser l'utilisation des ressources tout en limitant la consommation d'énergie, apportant ainsi une contribution significative à la durabilité

globale du réseau. Il convient de souligner que cette proposition soulève des défis techniques et nécessite une expertise appropriée pour garantir la validité et la fiabilité des résultats obtenus.

Ce mémoire comprend quatre chapitres :

- le premier chapitre fournit une vue détaillée des technologies LPWAN, mettant en évidence l'importance de ces réseaux dans le contexte de l'IoT. La technologie LoRaWAN, en particulier, a été examinée sous différents angles, démontrant son potentiel significatif dans divers secteurs. Ce chapitre sert de fondement pour une compréhension approfondie des enjeux et des opportunités liées aux technologies LPWAN dans le paysage complexe de l'IoT.
- le deuxième chapitre offre une exploration approfondie des techniques de clustering dans les réseaux IoT, couvrant les fondamentaux de l'apprentissage automatique, les types d'apprentissage, les caractéristiques du clustering, les méthodes et les algorithmes de clustering. La comparaison entre les différents algorithmes fournit une base solide pour choisir la méthode de clustering la plus adaptée en fonction des exigences spécifiques des réseaux IoT.
- le troisième chapitre examine en profondeur l'application de l'algorithme de clustering k-means dans le contexte des réseaux LoRa. L'utilisation de k-means dans LoRa offre des avantages significatifs pour la gestion des données IoT, mais il est crucial de prendre en compte les défis spécifiques liés à cette combinaison. La compréhension des limites des travaux existants et des problématiques soulevées dans l'article de référence contribue à éclairer le lecteur sur la complexité de cette intégration. Cette synthèse a pour objectif de fournir des connaissances précieuses pour les chercheurs et les praticiens travaillant dans le domaine des réseaux IoT et des technologies de clustering.
- le quatrième et le dernier chapitre présente une proposition novatrice qui est basée sur l'algorithme de clustering k-means dans le contexte spécifique des réseaux LoRa. L'amélioration de l'algorithme vise à minimiser les interférences Co-SF dans les réseaux IoT tout en prenant en compte les caractéristiques uniques de LoRa. L'analyse détaillée de la proposition et la discussion des algorithmes résultants fournissent une base solide pour évaluer l'efficacité de cette approche dans des scénarios pratiques. Cette contribution aspire à favoriser des développements significatifs dans ce domaine.

Chapitre 1 : Les technologies LPWAN dans l'IoT

Introduction

L'Internet des Objets (IoT) est une technologie qui transforme notre manière d'interagir avec le monde en connectant des milliards d'appareils intelligents pour recueillir et partager des données en temps réel. Cependant, pour que l'IoT fonctionne efficacement, une connectivité fiable et économique est indispensable. C'est là que les technologies LPWAN (Low Power Wide Area Network) entrent en jeu [3]. Les technologies LPWAN offrent une connectivité de longue portée et à faible consommation d'énergie, ce qui permet de connecter des appareils IoT à faible coût et sans besoin de changement fréquent des batteries. Dans ce chapitre, nous allons explorer les technologies LPWAN les plus populaires telles que LoRaWAN, Sigfox et NB-IoT, ainsi que leurs avantages, leurs limites et leurs applications dans divers secteurs.

I. Concept générale de l'IoT

L'Internet des Objets (IoT) représente un paradigme novateur dans le domaine des technologies de l'information et de la communication. Ce concept englobe la connectivité et l'interaction entre une multitude d'objets physiques, dotés de capteurs, d'actuateurs et de capacités de communication. L'objectif fondamental de l'IoT est de permettre aux objets de recueillir, de traiter et d'échanger des données de manière autonome, contribuant ainsi à une compréhension plus approfondie de l'environnement qui les entoure. Au cœur du concept de l'IoT réside la capacité des objets connectés à interagir entre eux et avec des systèmes informatiques, créant ainsi un écosystème dynamique [3]. Ces objets peuvent prendre diverses formes, des dispositifs portables aux équipements industriels, et sont souvent intégrés dans des réseaux sans fil pour assurer leur connectivité. L'IoT offre ainsi des possibilités considérables dans des domaines variés tels que la domotique, la santé, l'industrie, et bien d'autres, en facilitant la collecte de données en temps réel et en favorisant la prise de décision automatisée.

A. Architecture de réseau IoT

En général, l'architecture du réseau IoT se compose de deux couches : le réseau d'accès (RA) et le réseau fédérateur (RF). L'RA contient deux types de composants : les passerelles internet (point d'accès, station de base, NB, eNB, gNB [4], etc.) et une variété de dispositifs déployés. Les stations de base (SB) agissent comme une passerelle qui permet aux appareils du réseau de se connecter à l'internet. D'autre part, le NB se compose principalement d'Internet et envoie les

données collectées à des services distants spécifiques (pour la surveillance du réseau ou la sauvegarde et le traitement).

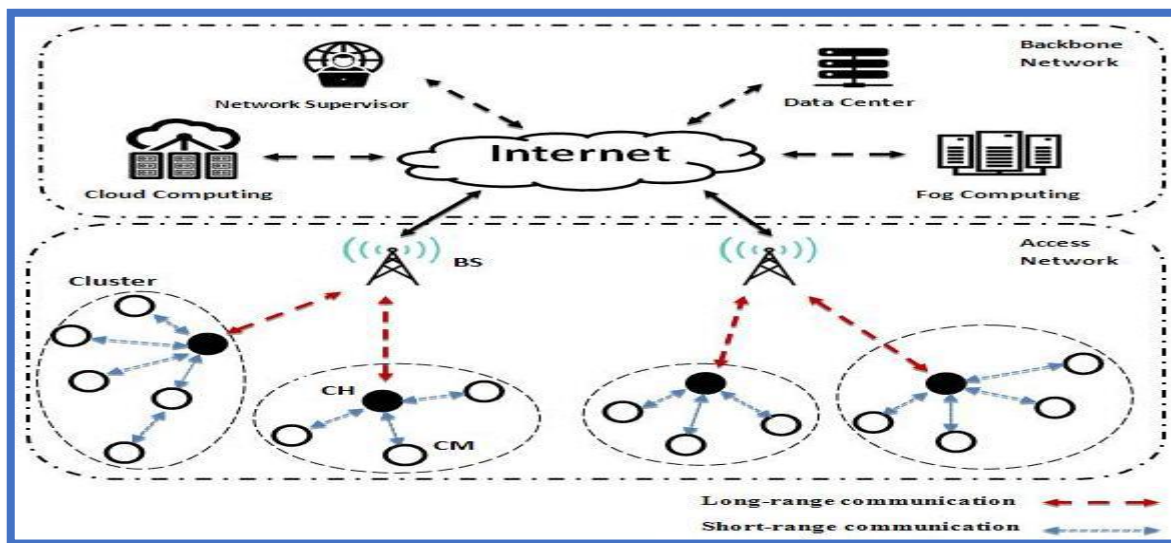


Figure 1: Architecture du réseau IoT [5]

B. Les technologies de communication de l'IoT

L'IoT est synonyme de connectivité et d'échange de données entre appareils/capteurs. En général, les applications IoT ont des exigences spécifiques telles qu'un faible débit de données échangées, une faible consommation d'énergie, une portée élevée et une rentabilité non négligeable. L'Internet des Objets (IoT) est alimenté par une diversité de technologies de communication qui répondent aux besoins spécifiques des applications variées dans des domaines tels que la domotique, la santé, l'industrie et bien d'autres. Parmi ces technologies, on distingue les solutions de courte portée, de longue portée bas débit et de longue portée haut débit, chacune offrant des avantages distincts en fonction des exigences particulières de chaque scénario.

1. Technologie de courte portée

Les technologies de courte portée sont conçues pour des communications sur de courtes distances, idéales pour les applications à proximité immédiate des dispositifs. Bluetooth, Wi-Fi[6], Zigbee, NFC et Z-Wave sont des exemples de technologies de courte portée utilisées dans l'IoT. La norme 802.15.4 est couramment utilisée à cette fin [7] et est adoptée par de nombreuses interface sans fil à faible cout telles que Bluetooth[8], Zigbee[9], WirelessHART[10], 6LoWPAN[11], Z-Wave[12], etc. La consommation d'énergie est réduite et les appareils peuvent fonctionner pendant des années sans être remplacés. Ces solutions sont

souvent privilégiées pour des interactions rapides, une faible consommation d'énergie et une connectivité directe entre des périphériques à proximité [3].

2. Technologie de longue portée haut débit

Les applications IoT nécessitant une couverture étendue tout en préservant une consommation d'énergie minimale adoptent des technologies de longue portée bas débit. Les solutions cellulaires (2G, 3G, 4G et bientôt 5G) peuvent fournir une plus grande couverture, mais elles consomment des quantités excessives d'énergie des appareils et entraînent des coûts de connectivité imposés par les opérateurs. La communication à longue portée présente une consommation d'énergie élevée et limite la durée de vie des dispositifs à quelques jours.

3. Technologie de longue portée bas débit

En outre, des technologies de communication avancées pour les zones étendues à faible consommation (LPWAN)[13] ont été étendues, telles que BLE (Bluetooth Low Energy)[14], WAVENIS[7], Wi-Fi (802.11), LTE. Le choix de l'interface de communication dépend de l'application du système. Ainsi les besoins des applications IoT ont conduit à l'apparition d'une nouvelle technologie de communication sans fil : Low Power Wide Area Network ou LPWAN qu'il est nécessaire de prendre maintenant en compte comme candidat sérieux aux réseaux dédiés à l'IoT. LoRaWAN (LoRa)[15] et Narrowband IoT (NB-IoT). Le tableau 1 montre les technologies de communications existant et leurs applications associées.

Protocole	Bande de Fréquences	Taux de données	Gamme	Consommation d'énergie	Débit
Wi-Fi	2.4/5.0 GHz	54 Mb/s	30 m	Mayen	Courte
Bluetooth	2.4GHz	1 Mb/s	10 m	Faible	Courte
IEEE 802.15.4	868-915 MHz/2.4 GHz	250 kb/s	[10,100] m	Faible	Bas
LTE	2.6 GHz	10 Mb/s	15 km	Haut	Bas

HART sans fil	2.4-2.5 MHz	250 kb/s	[1,100] m	Bas	Bas
BLE	2.4 GHz	1 Mb/s	200 m	Très faible	Bas
Z-WAVE	1 GHz	40 kb/s	30 m	Faible	Courte
WAVENIS	865-916 GHz	100 kb/s	4 km	Très faible	Bas
LoRaWAN	868-900 GHz	50 kb/s	15 km	Très faible	Bas
NB-IoT	180 kHz	234.7 kb/s	35 km	Faible	Bas

Tableau 1 : Comparaison des technologies de communication de l’IoT

Chaque protocole présente des caractéristiques spécifiques en termes de bande de fréquences, taux de données, portée, consommation d'énergie et débit, adaptées à des cas d'utilisation particuliers dans le domaine de l'IoT. Le choix du protocole dépend des exigences spécifiques de chaque application, en fonction de facteurs tels que la portée requise, la consommation d'énergie, et le débit nécessaire.

II. Les Technologies LPWAN

Les dispositifs IoT sont généralement alimentés par une batterie pour permettre la mobilité. Pour cette raison, le profil de consommation d'énergie doit être soigneusement étudié afin d'étendre la durée de vie de la batterie. La portée de communication doit atteindre plusieurs kilomètres, car les appareils finaux sont répartis sur une vaste zone, comme les bâtiments et les champs agricoles. De nombreuses technologies LPWAN sont déjà disponibles, comme Sigfox[16], Narrow Band-Internet of Things (NB-IoT)[17] ou LoRaWAN[18]. Sigfox prévoit

d'offrir une couverture mondiale dans 45 pays et régions par une entreprise de télécommunication comme une alternative aux technologies LPWAN sub-GHz. Comme NB-IoT utilise un spectre sous licence, il offre une meilleure fiabilité du trafic par rapport aux autres technologies sub-GHz.

Ce pendant LoRaWAN pourrait être déployé en tant que réseau privé et s'intégrer facilement à de nombreuses plateformes de réseau (par exemple, The Things Network (TTN)). Abordons maintenant les caractéristiques fondamentales des technologies LPWAN, afin de mieux saisir leur impact et leurs avantages dans le domaine de l'Internet des Objets.

A. Les caractéristiques des technologies LPWAN

Le succès des technologies LPWAN réside dans leur capacité à offrir une connectivité à faible consommation à un nombre considérable de terminaux, répartis sur de grandes zones géographiques, à faible coût. Cette section décrit les techniques utilisées par les technologies LPWAN pour atteindre ces objectifs souvent contradictoires [15]. Nous soulignons que les technologies LPWAN partagent certains des objectifs de conception avec d'autres technologies sans fil. L'objectif principal des technologies LPWAN est toutefois d'atteindre une large portée avec une faible consommation d'énergie et un faible coût, contrairement aux autres technologies pour lesquelles il peut être plus important d'obtenir un débit plus élevé, une latence réduite et une fiabilité accrue.

1. La portée

Les technologies LPWAN sont conçues pour couvrir une large zone et permettre une bonne propagation du signal vers des endroits intérieurs difficiles d'accès tels que les sous-sols. Sur le plan quantitatif, un gain de +20 dB par rapport aux systèmes cellulaires existants est visé. Cela permet aux terminaux de se connecter aux passerelles à une distance de quelques dizaines de kilomètres, en fonction de l'environnement de déploiement (rural, urbain, etc.) [17]. La bande sous 1 GHz et les schémas de modulation spéciaux sont exploités pour atteindre cet objectif.

Différentes variantes des techniques d'étalement du spectre sont utilisées par les normes existantes. Le spectre d'étalement de propagation (CSS) et le spectre d'étalement par séquence directe (DSSS) sont utilisés respectivement par LoRa et RPMA.

2. La consommation d'énergie

Le fonctionnement à très faible consommation d'énergie est une condition essentielle pour tirer parti de l'énorme opportunité commerciale offerte par les terminaux IoT alimentés par batterie. Une durée de vie de la batterie de 10 ans ou plus avec des piles AA ou à pile bouton est souhaitable pour réduire les coûts de maintenance. Une basse consommation dépend des paramètres suivants : la topologie[21], le cycle de service, le protocole d'accès au médium simplifié [22] et la simplification de la complexité des terminaux. La topologie résultante LoRa est une étoile (b) qui est largement utilisée dans les réseaux cellulaires et apporte un avantage en matière d'économie d'énergie. D'autre part il existe des topologies maillées (a) et arborescentes (c). Le cycle de service permet aux terminaux LPWAN d'éteindre leurs émetteurs-récepteurs lorsqu'ils ne sont pas nécessaires, ils sont activés uniquement lorsque les données doivent être transmises ou reçues. Avec le grand nombre de terminaux, à priori les technologies LPWAN ne peuvent pas supporter cette surcharge de signalisation excessive. Néanmoins, INGENU et NB-IoT considèrent également que les protocoles MAC basés sur le TDMA permettent d'allouer les ressources radio plus efficacement, bien qu'au prix d'une complexité et d'un coût accrus pour les terminaux. Nonobstant, des recherches se produisent pour appliquer le TDMA au LoRaWAN. LoRaWAN est un exemple notable dans lequel le cœur du réseau adapte les paramètres de communication (tels que les paramètres de débit / modulation) pour maintenir de bonnes connexions en liaisons montante et descendante [23].

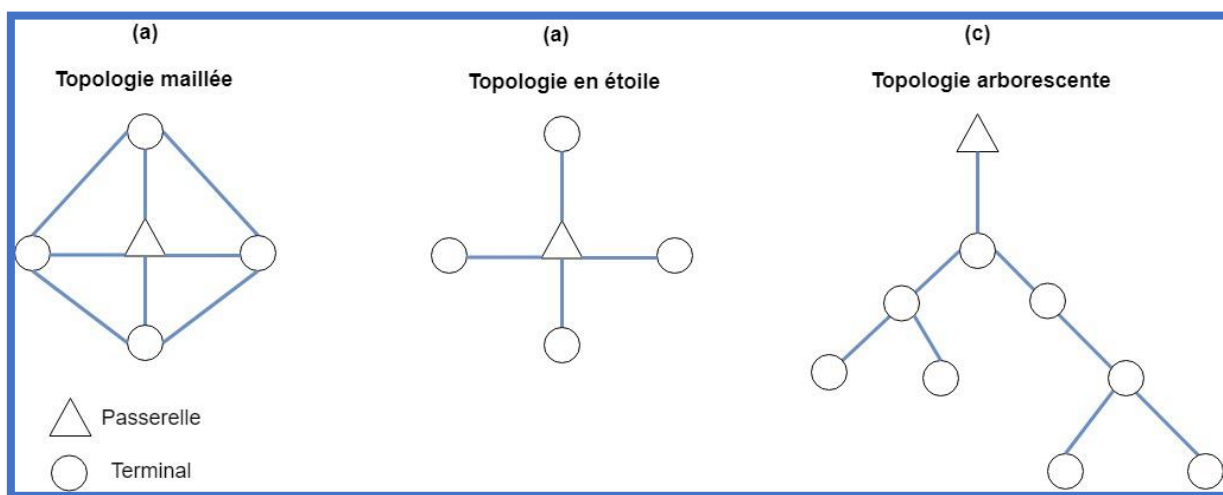


Figure 2: Topologies en étoile, maillée et en arbre [24]

3. Les couts

Le succès commercial des réseaux LPWAN est lié à la connexion d'un grand nombre de terminaux, tout en maintenant le coût du matériel en dessous de 5\$ et l'abonnement de connectivité par unité à 1\$. Cette abordabilité permet aux technologies LPWAN non seulement de s'adresser à un large éventail d'applications, mais aussi de concurrencer favorablement dans les domaines où les technologies sans fil à courte portée et les réseaux cellulaires sont déjà bien établis. Les technologies LPWAN adoptent plusieurs méthodes pour réduire les CAPEX et les OPEX pour les utilisateurs et les opérateurs de réseau. La conception à faible coût des terminaux est rendue possible par plusieurs techniques, dont certaines ont été déjà décrites précédemment. L'utilisation d'une connectivité de type en étoile (au lieu d'un maillage), de simples protocoles MAC et de techniques permettant de réduire la complexité des terminaux permet aux fabricants de concevoir des terminaux simples et donc économiques [24].

4. L'évolutivité

La prise en charge d'un grand nombre de terminaux échangeant des faibles volumes de trafic est l'une des principales exigences des technologies LPWAN. Ces technologies doivent donc s'adapter au nombre et à la densité croissante de terminaux connectés. Plusieurs techniques sont envisagées pour faire face à cette contrainte : la diversité, la densification, la sélection de canal adaptatif et débit de données.

Passons maintenant à examiner de plus près trois des technologies LPWAN les plus utilisées : Sigfox, NB-IoT et LoRa, pour voir comment ces caractéristiques se traduisent dans leur conception et leur fonctionnement respectifs.

B. Sigfox

SigFox est sans doute la société qui a le plus d'influence dans le domaine des réseaux LPWAN grâce à ses campagnes de marketing réussies en Europe. Il bénéficie également d'un excellent écosystème de fournisseurs, dont Texas Instruments, Silicon Labs et Axom. Sigfox est une technologie simple et complète qui utilise un schéma de modulation par déplacement de phase binaire (BPSK) pour envoyer les données. Il utilise une bande ultra étroite de 100 Hz, envoyant de très petites données (12 octets) très lentement (300 bauds) [3]. SigFox est un exemple d'utilisation d'un taux de modulation lent pour obtenir une plus grande portée. Cependant, il présente également des limites majeures : les capacités de liaison descendante sont fortement limitées et les collisions deviennent un problème majeur.

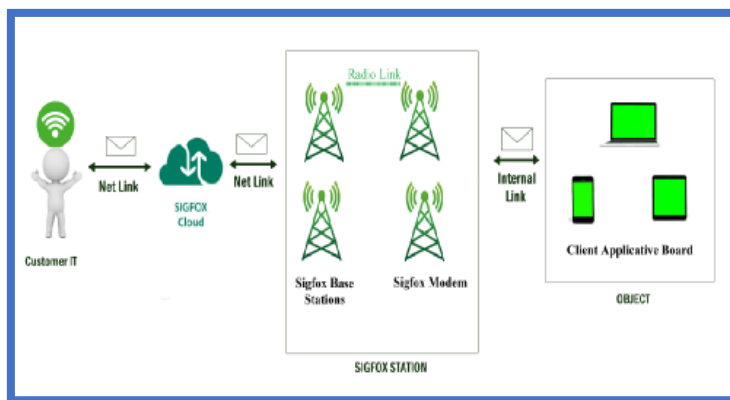


Figure 3: Architecture de Sigfox [21]

C. Narrow Band-Internet of Things (NB-IoT)

NB-IoT peut être déployé en réutilisant et en mettant à niveau le réseau cellulaire existant, mais ses déploiements sont limités à la zone prise en charge par le réseau cellulaire. La spécification NB-IoT a été publiée en juin 2016, et il faudra donc un délai supplémentaire pour établir le réseau NB-IoT. D'autre part, les composants LoRa et l'écosystème LoRaWAN sont matures et prêts pour la production maintenant, bien que les déploiements à l'échelle nationale soient encore en phase de déploiement [19].

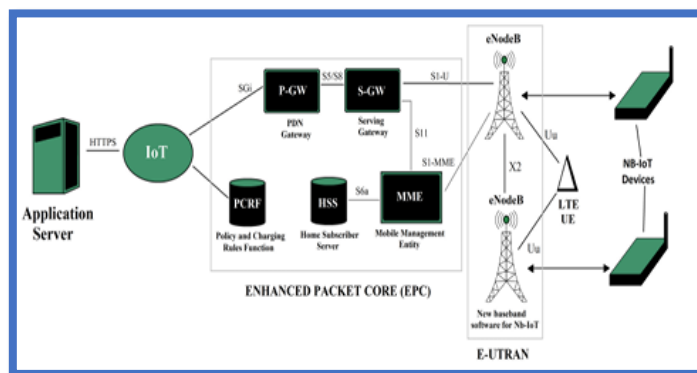


Figure 4: Architecture de NB-IoT [20]

D. LoRaWAN

LoRaWAN utilise un spectre sans licence et est un protocole asynchrone. LoRa, basé sur la modulation CSS, peut gérer les interférences, les trajets multiples et les évanouissements, mais il ne peut pas offrir la même qualité de service que NB-IoT [23]. En effet, NB-IoT utilise un spectre sous licence et son protocole synchrone à créneaux temporels est optimal pour la QoS. Toutefois, cet avantage de la QoS se fait au détriment du coût. Les enchères de spectre sous licence pour le spectre sub-GHz sont généralement supérieures à 500 millions de dollars par

MHz [8]. En raison du compromis entre la QoS et le coût élevé du spectre, les applications qui ont besoin de QoS préfèrent le NB-IoT, tandis que les applications qui n'ont pas besoin de QoS élevée devraient choisir LoRa.

Cependant la comparaison de ces technologies peut certainement offrir une meilleure compréhension de leurs différences et de leurs avantages respectifs.

E. Comparaison des technologies LPWAN

Les cas d'étude de l'IoT sont largement classés en quatre catégories, à savoir l'IoT personnel, l'IoT public, l'IoT industriel et les appareils IoT. Chaque catégorie est mieux identifiée comme NB-IoT ou LoRa. LoRa, tout comme NB-IoT, représente 45 % du marché commercial des LPWAN [24]. Le champ LPWA LoRa est vaste, avec des communications allant de quelques mètres à plus de 100 km. La rentabilité du réseau NB-IoT permet d'encadrer un grand nombre d'appareils dont la durée de vie de la batterie est supérieure à 10 ans [25]. Le LoRa utilise un spectre sans licence et est un protocole asynchrone. Un tableau récapitulatif des technologies LPWAN (Low Power Wide Area Network) peut inclure plusieurs paramètres pertinents pour comparer ces technologies.

Technologie LPWAN	Portée maximale	Débit de données	Consommation d'énergie	Pénétration du bâtiment	Fréquence de fonctionnement
LoRa (Long Range)	2-15 km	0.3 kbps - 50 kbps	Très Faible	Excellente	868 MHz (Europe), 915 MHz (Amérique du Nord), 433 MHz (Asie)
Sigfox	30-50 km	Jusqu'à 100 bps	Faible	Bonne	868 MHz (Europe), 902 MHz (Amérique du Nord), 920 MHz (Asie)
NB-IoT (Narrowband IoT)	10 km	20 kbps - 250 kbps	Faible	Bonne	Bandes sous licence dans les spectres cellulaires existants

Tableau 2 : Tableau récapitulatif des technologies LPWAN

D'après le tableau, LoRa offre une portée significative, ce qui en fait un choix solide pour les applications nécessitant une couverture étendue. Les débits de données modérés conviennent aux applications où la transmission de petites quantités de données à des intervalles plus longs est acceptable. Mais aussi la Consommation d'énergie faible : Idéal pour les dispositifs alimentés par batterie. La plupart des commandes de puissance de transmission visent à augmenter la capacité, et pas seulement à réduire la consommation d'énergie. Le contrôle de la vitesse de transmission ne vise souvent qu'à maximiser le débit. Par rapport à LoRa, les taux de paquets Wifi sont nettement plus élevés, et les algorithmes de contrôle du débit d'émission fonctionnent à un rythme beaucoup plus élevé que dans LoRa. En outre la spécification LoRaWAN est en accès libre. En tenant compte de ces paramètres, ainsi que des exigences spécifiques de votre application IoT, choisir LoRa peut offrir une solution robuste et rentable pour connecter des dispositifs sur de longues distances avec une consommation d'énergie minimale.

III. La technologie LoRaWAN

LoRaWAN est une nouvelle technologie de communication à longue portée proposée par Semtech qui repose sur une modulation par étalement de spectre chirp qui utilise toute la bande de fréquence pour moduler des impulsions chirp. Un chirp est un signal sinusoïdal dont la fréquence augmente ou diminue au fil du temps et qui code un certain nombre de bits d'information. Contrairement à la modulation FSK la plus courante, la modulation LoRa[26] mais améliore l'immunité au bruit et aux interférences et, par conséquent, augmente la portée des communications. Le résultat est qu'une seule gateway (GW) peut couvrir une région de plusieurs kilomètres carrés. Alors que LoRa définit la couche physique et est une technologie propriétaire (Semtech 2015), la spécification est disponible publiquement et elle est promue par l'opensource (LoRa Alliance Technical Committee 2017). Comme le montre la figure une architecture LoRaWAN est basée sur trois composants principaux :

- ✓ Ed : est le capteur/actionneur à faible consommation d'énergie qui communique avec les GW en utilisant la modulation LoRa.
- ✓ GW : élément intermédiaire qui collecte les paquets provenant des ED et les transmet au NS via un backaul IP (par exemple, Ethernet, 3G). Il peut y avoir plusieurs GW dans un déploiement LoRa.

- ✓ NS : est le serveur réseau responsable de la déduplication et décodage des paquets envoyés par les EDs. Les informations relatives aux paquets sont envoyées au serveur d'application. Le NS peut également générer des paquets à renvoyer aux EDs, lorsqu'une configuration ED est nécessaire [27].

Le réseau LoRaWAN présente une topologie en étoile et, contrairement aux réseaux cellulaires traditionnels, les DE ne sont pas associés à un GW spécifique. LoRaWAN ne permet pas les communications de dispositif à dispositif, les paquets ne peuvent être transmis d'un ED au NS, ou vice-versa.

Dans le système LoRa, les ED prennent en charge les débits multiples en exploitant six facteurs d'étalements (SF) différent de 7 à 12. Le choix du SF a un impact sur la durée et la probabilité du paquet généré. Les communications sur différentes SF dans le même canal sont en principe orthogonales (Croce et al. 2017).

Maintenant que nous avons passé en revue les caractéristiques fondamentales de la technologie LoRa, examinons de plus près ses différents aspects, notamment son architecture LoRaWAN, les dispositifs finaux, la qualité de service (QoS) qu'elle offre, son application dans le secteur professionnel, ainsi qu'un aperçu de cette technologie.

A. Architecture LoRaWAN

L'architecture du réseau LoRaWAN comprend des dispositifs terminaux, des serveurs de réseau et des passerelles LoRa, comme le montre [28] la figure 5. La connexion entre les passerelles et le serveur est fiable et rapide, et les passerelles sont connectées aux capteurs sans fil à l'aide de la technologie LoRa. Le serveur est le coordinateur du réseau et les passerelles jouent le rôle de répéteurs entre les capteurs et le serveur. Après avoir reçu une trame via une connexion sans fil ; la passerelle encapsule la trame dans un paquet IP et la transmet au serveur et, de la même manière, transmet les paquets du serveur aux capteurs.

LoRaWAN utilise une architecture en étoile à longue portée dans laquelle une passerelle est utilisé pour relayer les messages entre les dispositifs terminaux et un réseau central. Dans un réseau LoRaWAN, les nœuds ne sont pas associés à une passerelle spécifique. Au contraire, les données transmises par un nœud sont généralement reçues par plusieurs passerelles. Chaque passerelle transmet le paquet reçu du nœud final au serveur du réseau en nuage via une liaison de retour (cellulaire, Ethernet, satellite ou Wi-Fi) [29].

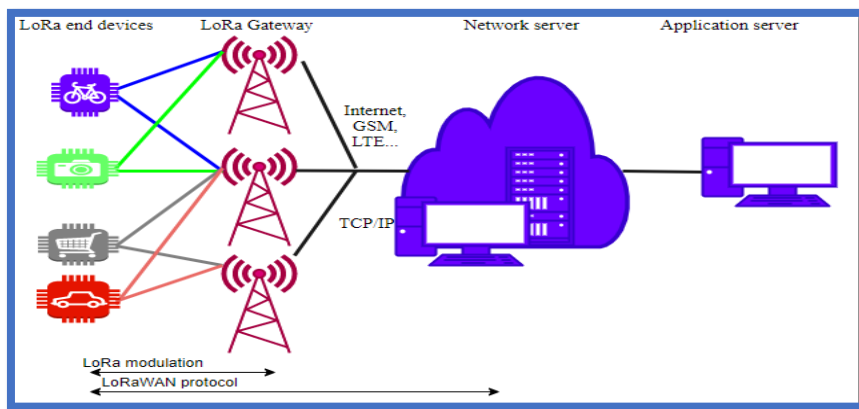


Figure 5: Architecture du réseau LoRaWAN

B. Les dispositifs finaux

Les dispositifs finaux LoRa ou end device en anglais sont des appareils qui utilisent la technologie de communication sans fil LoRa pour transmettre des données sur de longues distances avec une consommation d'énergie minimale. Conçus pour être efficaces en termes de consommation énergétique et de portée de transmission, ces dispositifs sont largement utilisés dans diverses applications telles que l'Internet des objets (IoT), les réseaux de capteurs, la surveillance environnementale, la gestion intelligente des ressources [29].

Dans le contexte de l'IoT, les dispositifs finaux peuvent être des capteurs, des actionneurs, des smartphones, des ordinateurs, des équipements de surveillance, des véhicules connectés, etc. Voici un exemple de dispositif final qui est le Capteur de Qualité des Sols LoRaWAN.



Figure 6 : Capteur de Qualité des sols LoRaWAN

Le capteur de qualité des sols LoRaWAN est un dispositif compact et autonome qui intègre plusieurs capteurs spécialisés pour mesurer divers paramètres clés liés à la qualité du sol. Il est

conçu pour être facilement déployé dans les champs agricoles et d'autres environnements. Le dispositif est équipé de capteurs spécialisés pour mesurer des paramètres tels que la teneur en humidité du sol, la température du sol, le pH du sol, les niveaux de nutriments (azote, phosphore, potassium), la conductivité électrique du sol, ainsi que d'autres indicateurs de la santé et de la fertilité du sol.

C. Les interférences dans LoRa

Les interférences dans LoRa (Long Range) peuvent être causées par divers facteurs tels que d'autres dispositifs sans fil fonctionnant sur la même fréquence, des obstacles physiques, des conditions météorologiques, ou même des interférences intentionnelles et elles peuvent avoir un impact significatif sur la qualité des communications[29]. Nous allons expliquer dans la suite les principaux points à retenir concernant les interférences dans LoRa :

Interférences de bande passante : Les transmissions non LoRa dans la même bande de fréquence peuvent créer des interférences et perturber les communications LoRa. Il est important de choisir des fréquences de fonctionnement appropriées pour minimiser ce type d'interférences.

Interférences physiques : Des obstacles physiques tels que des bâtiments, des arbres denses, ou d'autres éléments du paysage peuvent affaiblir le signal LoRa ou entraîner des réflexions pouvant perturber la communication entre les nœuds du réseau.

Interférences atmosphériques : Les phénomènes météorologiques tels que les précipitations, la neige ou la foudre peuvent influencer la propagation des ondes radio et provoquer des interférences temporaires dans les communications LoRa.

Interférences provenant d'autres dispositifs : La présence d'autres dispositifs sans fil opérant dans la même bande de fréquence peut causer des interférences avec les transmissions LoRa. Il est crucial de prendre en compte ces sources potentielles d'interférences lors de la planification du réseau.

En prenant en compte ces facteurs et en mettant en œuvre des mesures pour atténuer les interférences, comme le positionnement approprié des antennes, la sélection judicieuse des canaux de fréquence et la surveillance régulière de la qualité du signal, il est possible d'optimiser les performances et la fiabilité des réseaux LoRa en limitant les impacts des interférences.

Les facteurs d'étalement sont des paramètres clés dans la modulation par étalement de spectre, une technique utilisée dans des systèmes comme LoRa pour améliorer la robustesse du signal

et résister aux interférences[26]. Il est important de noter que des dispositifs LoRa ayant le même facteur d'étalement peuvent entraîner des interférences s'ils fonctionnent dans la même gamme de fréquences et se chevauchent les uns les autres. Même si deux dispositifs utilisent des facteurs d'étalement identiques, ils peuvent toujours perturber les communications l'un de l'autre s'ils émettent simultanément et se trouvent dans la zone de couverture respective.

D. La qualité de services (QoS) dans LoRa

Étant donné que la nécessité d'une QoS élevée ou réduite dépend fortement des exigences des applications fonctionnant sur les équipements terminaux, les serveurs de réseau LoRaWAN doivent pouvoir affecter des paramètres de modulation LoRa basés sur les valeurs d'appartenance aux exigences de l'application. Par exemple, si l'application s'exécute sur les dispositifs terminaux doit envoyer des requêtes de paquets de données avec une QoS élevée, le serveur de réseau doit sélectionner les paramètres requis dans un pool ou un groupe de paramètres avec RSSI élevé, lowToA et lowBER. Les serveurs de réseau LoRaWAN devraient être capables de choisir les paramètres de transmission en fonction de leur appartenance à différents types d'applications. Des courts délais de retard limitent généralement la capacité des terminaux à passer en mode veille et réduisent donc la durée de vie de la batterie [30]. La meilleure approche consiste à mettre en place un système flexible capable d'obtenir différents résultats en termes de QoS en fonction des besoins, en privilégiant le trafic à haute QoS (comme les alarmes) par rapport au trafic le plus bas.

E. La technologie LoRaWAN dans le secteur professionnel

L'utilisation de la technologie LoRaWAN est très variée dans le monde. Elle peut toucher tout autant le secteur du bâtiment, de capteurs sur les lieux publics. Elle peut aussi être utilisée dans le secteur agricole ou voir même dans le contexte de la Covid-19 pour notamment les traqueurs de proximité.

1. Bâtiment intelligent

Dans le secteur du Smart Building ou bâtiment intelligent, il y a plusieurs cas où des appareils LoRa peuvent être utiles à implémenter. Dans le cadre d'un bâtiment universitaire, ces capteurs serviraient par exemple de vérifier si une classe est actuellement occupée et pourrait envoyer l'information à un serveur gérant tous ces appareils [27]. Nous aurons ainsi la possibilité d'avoir une vision globale de l'utilisation de chaque salle. Des capteurs de mouvements peuvent servir

à analyser le flux de passage dans le bâtiment, qui pourrait être utile dans le cadre d'une pandémie tel que le Covid-19.

2. Agriculture

Dans le domaine de l'agriculture, les appareils LoRa peuvent être utilisés comme capteurs d'environnement. La portée élevée des appareils permet un répartition plus facile sur un grand terrain tout en gardant la communication avec le serveur LoRaWAN. Les mesures de débits d'eau ou d'irrigation sont de petites données et peuvent donc être transmis par LoRa. LoRa peut aussi être utilisé dans le traçage de positions des bétails. En effet, des capteurs de positions placés sur des vaches peuvent transmettre les positions à intervalle régulier vers un serveur LoRaWAN. Il devient alors intéressant dans des régions où les connectivités sont peu présentées du bétail.

3. Ville intelligente

Le smart city est aussi un marché en plein essor et où il y aurait une nécessité de suivi essor dans les grandes villes. En effet, ces dernières déploient de plus en plus de capteurs dans la ville. Il existe par exemple des capteurs de détection de véhicule. L'entreprise française IEM propose notamment PrestoSense qui est un capteur qui peut être posé au sol des places de parking en ville et ainsi connaître en temps réel leur occupation. Nous pouvons voir en temps réel l'utilisation des parkings publics dans certaines rues équipées de ces appareils. Il y a aussi la possibilité d'utiliser des capteurs de la qualité d'air avec une transmission LoRa.

F. Aperçu de la technologie LoRa

La technique propriétaire de la couche physique LoRa est détenue par Semtech Corporation. L'Alliance LoRa a spécifié LoRaWAN comme protocole de la couche de contrôle d'accès au support (MAC). Cette partie décrit plus précisément la technologie LoRa, LoRaWAN et ses principales caractéristiques.

1. Couche physique LoRa

La modulation LoRa comporte plusieurs paramètres :

- Facteur d'étalement (SF) ;
- Largeur de bande (BW) ;
- Spectre d'étalement chirp (CSS) ;
- Taux de code (CR) [26].

Dans la modulation LoRa, les informations sont transmises sous formes de symboles, dont la longueur T_s dépend de l'utilisation de (SF). Chaque symbole est un signal sinusoïdal dont la fréquence est décalée cycliquement dans une largeur de bande (BW). La modulation LoRa comprend également un système de correction d'erreur variable qui améliore la robustesse de la transmission du signal au détriment de la redondance. Un autre paramètre critique est la sensibilité du récepteur, qui indique le niveau de puissance le plus faible du signal LoRa reçu que le récepteur peut détecter et démoduler. Le tableau 3 indique le débit binaire nominal et la sensibilité du récepteur pour la bande passante de 125 khz. Pour les valeurs de tableau 1, la portée maximale de communication du LoRa est d'environ 10 km [2,11].

Bande passante (kuz)	SF	Débit nominal Rb (bps)	Sensibilité (dBm)
125	6	9375	-118
125	7	5469	123
125	8	3125	126
125	9	1758	129
125	10	977	132
125	11	537	134
125	12	293	137

Tableau 3 : Vitesse de transmission et sensibilité du récepteur à partir du facteur d'étalement (SF)[2,11]

2. Couche MAC LoRaWAN

LoRaWAN est un protocole conçu pour de nombreuses applications LPWAN qui utilisent des bandes de fréquences sans licence pour la transmission. Sa norme a été publiée en 2005 et décrit le protocole de la couche liaison de données, tandis que le protocole de la couche physique est propriétaire et appartient au fabricant de l'émetteur. La figure 7 montre le protocole défini de LoRaWAN via l'Alliance LoRa [32].

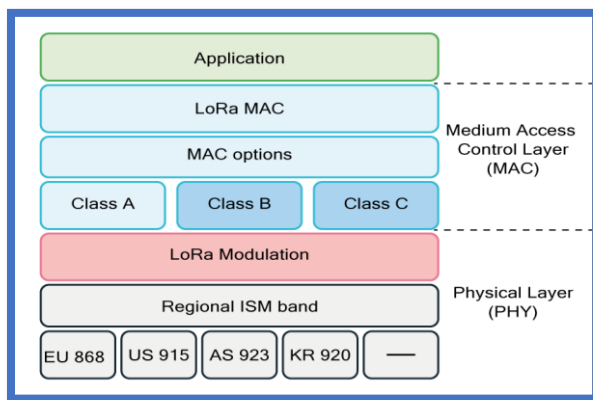


Figure 7: Pile de LoRaWAN

Les dispositifs LoRaWAN sont classés en trois classes : A, B et C. La classe A est la principale, basée sur la méthode d'accès ALOHA asynchrone [33], et doit être prise en charge par tous les dispositifs. Tous les capteurs LoRaWAN, lorsqu'ils sont allumés, fonctionnent selon la classe A et peuvent passer à d'autres classes si une telle possibilité physique est disponible et après accord avec le serveur. La classe B repose sur la distribution périodique d'information sur les services par le serveur et sur l'accès au canal selon un calendrier. Cela suppose que les dispositifs peuvent consommer plus d'énergie que les dispositifs de classe A, ce qui leur permettra d'écouter les messages périodiques du serveur. La classe C est basée sur l'écoute constante du canal par les capteurs.

Conclusion

Les technologies LPWAN (Low Power Wide Area Network) ont un potentiel considérable pour l'IoT (Internet des objets), car elles permettent de connecter des milliards d'appareils intelligents de manière économique, efficace et à longue portée. Les principales technologies LPWAN telles que LoRaWAN, Sigfox et NB-IoT offrent des avantages différents en termes de portée, de débit de données et de consommation d'énergie, ce qui permet aux utilisateurs de choisir la meilleure solution en fonction de leurs besoins spécifiques.

Cependant, il est important de noter que les technologies LPWAN ont également des limites, telles que la capacité de bande passante limitée, la latence élevée et la vulnérabilité à des interférences. Par conséquent, il est important de considérer ces facteurs lors de la conception et de la mise en œuvre de systèmes IoT utilisant des technologies LPWAN.

Malgré ces limites, les technologies LPWAN ont déjà été largement adoptées dans divers secteurs, tels que l'agriculture intelligente, la ville intelligente, la surveillance de

l'environnement et l'industrie 4.0. À mesure que les avantages des technologies LPWAN deviennent plus évidents, nous pouvons nous attendre à ce que leur adoption continue de croître et à ce que de nouveaux cas d'utilisation émergent.

Après avoir exploré en profondeur les technologies LPWAN dans l'IoT, nous nous tournons désormais vers l'application des techniques de clustering dans les réseaux IoT, dévoilant ainsi les stratégies de gestion et d'analyse des données massives générées par ces dispositifs interconnectés.

Chapitre 2 : Les techniques de clustering dans les réseaux IoT

Introduction

L'Internet des objets (IoT) a permis la création de réseaux massifs d'objets interconnectés. Ces objets collectent des données dans divers domaines, tels que l'agriculture, l'industrie et la santé, offrant ainsi des avantages considérables dans ces secteurs. Cependant, la gestion et le traitement de ces vastes quantités de données peuvent être difficiles. C'est là que les techniques de clustering dans les réseaux IoT entrent en jeu. Le clustering est une technique de regroupement des objets IoT en clusters pour faciliter la gestion et l'analyse des données. Les techniques de clustering permettent d'optimiser l'efficacité énergétique, d'améliorer la qualité de service, de prolonger la durée de vie des batteries des objets et de réduire les coûts de communication dans les réseaux IoT. Dans ce chapitre, nous faisons un petit rappel sur l'apprentissage automatique et explorerons les différentes techniques de clustering utilisées dans les réseaux IoT, leurs avantages, leurs limites et leurs applications dans divers domaines.

I. Apprentissage automatique

Le terme apprentissage automatique a été inventé chez IBM en 1959 par Arthur, un pionnier américain des jeux informatiques et de l'IA, en tant que projet scientifique.

L'apprentissage automatique est la programmation d'ordinateurs pour optimiser les critères de performance à l'aide de données d'échantillon ou d'expériences passées. L'apprentissage est le processus de définition d'un modèle jusqu'à certains paramètres et d'exécution d'un programme informatique qui utilise des données de formation et l'expérience passée pour optimiser les paramètres du modèle. Le modèle peut être prédictif pour faire des prédictions dans la futur, ou descriptif pour acquérir des connaissances des données, ou les deux. L'objectif visé est de rendre la machine ou l'ordinateur capable d'apporter des solutions à des problèmes complexes par le traitement d'une quantité importante d'informations [35]. Par conséquent, il fournit un moyen d'analyser et de mettre en évidence les corrélations qui existent entre deux ou plusieurs situations spécifiques et de prédire leurs divers effets. Ci-dessous, nous définissons les concepts de base de l'apprentissage automatique.

II. Les types d'apprentissage automatique

Des milliers d'applications utilisent l'apprentissage automatique, notamment Facebook, Netflix, Google Translate et d'autres applications médicales...etc.

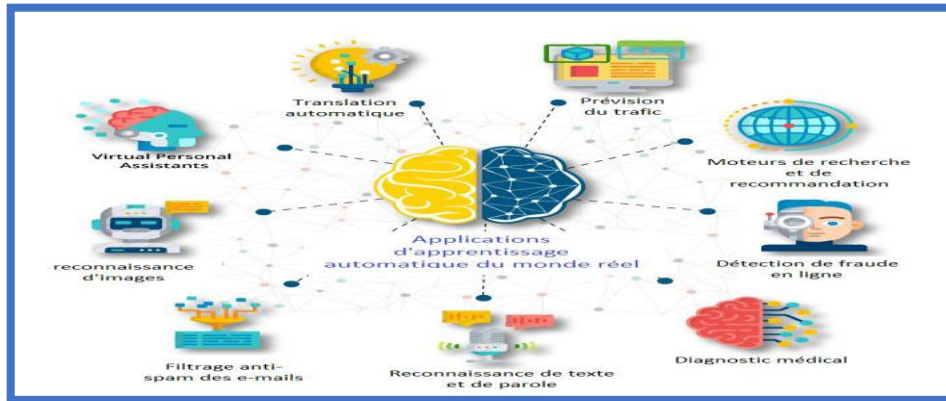


Figure 8: Les applications les plus importantes de l'apprentissage automatique [34]

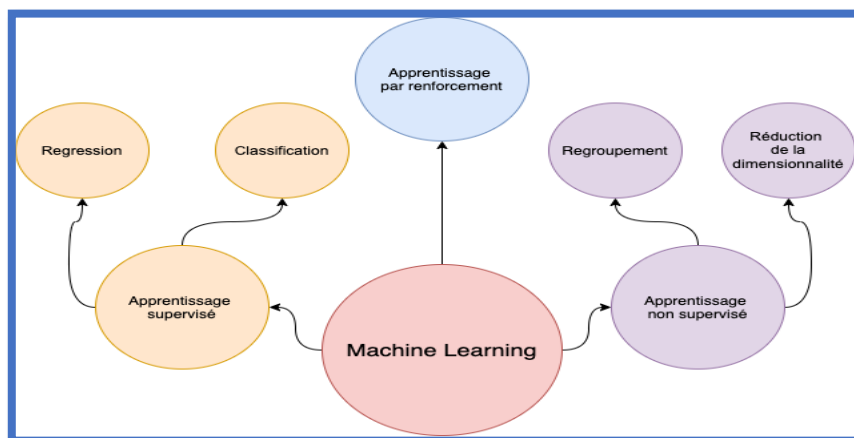


Figure 9: Les types d'apprentissage automatique [34]

A. Apprentissage supervisé

L'apprentissage supervisé consiste à former un modèle avec des étiquettes de réponse. Cette réponse vous aide à surveiller l'apprentissage de votre modèle en lui indiquant à quel point vous êtes loin de la bonne réponse. Dans un apprentissage supervisé, nous avons un X (variable indépendante) et un Y (variable dépendante) lors de l'entraînement. Cette catégorie se divise en 2 sous catégories principales soit la classification et la régression [36].

1. Classification

La classification est le processus de catégorisation d'un ensemble de données donné en classes. Cela peut être fait avec des données structurées ou non structurées. Le processus commence par prédire la classe de points de données. Les classes sont souvent appelées cibles, étiquettes ou catégories.

2. Régressions

L'analyse de régression est un concept fondamental de l'apprentissage automatique. Celle-ci est classé comme apprentissage supervisé où l'algorithme est formé à la fois sur les caractéristiques d'entrée et les étiquettes de sortie. Elle aide à établir une relation entre les variables en estimant comment une variable affecte l'autre [37].

Plusieurs algorithmes sont basés sur ce type d'apprentissage, parmi eux on peut citer : Régression linéaire, Régression logistique, Arbres de décision, K-NN, Machines à vecteurs de support, Random Forest...

B. Apprentissage non supervisé

L'apprentissage non supervisé consiste à trouver des caractéristiques et à extraire des relations entre les données. Dans les problèmes non supervisés, il n'y a pas de réponse exacte à trouver par le modèle. Vous n'avez que des données d'entrée et le modèle est en quelque sorte « laissé à lui-même ». En apprentissage non supervisé, nous avons seulement un X (variable indépendante) lors l'entraînement. L'apprentissage non supervisé se divise lui aussi en 2 principales catégories soient le regroupement (clustering) et la réduction de la dimensionnalité (dimensionnalité réduction).

Ce type d'apprentissage contient l'algorithme suivant : k-means, Fuzzy C-Means

1. Clustering

Le clustering est le processus de division d'une population ou de points de données en groupes afin que les points de données d'un même groupe soient plus similaires aux autres points de données du même groupe que les points de données d'autres groupes. En termes simples, l'objectif est de séparer les groupes ayant des traits similaires et de les affecter en clusters [38].

2. La réduction de la dimensionnalité

La réduction de la dimensionnalité est également réalisée dans le cadre de l'apprentissage non supervisé, qui est une branche de l'apprentissage automatique dans laquelle l'algorithme ne

reçoit pas d'étiquettes ou de réponses préalables. Dans ce contexte, la réduction de la dimensionnalité est souvent utilisée pour extraire des caractéristiques significatives à partir de données brutes, afin de faciliter leur analyse et leur compréhension.

C. Apprentissage par renforcement

L'apprentissage par renforcement consiste en des agents interagissant avec l'environnement. Ceci est un schéma montrant comment fonctionne l'apprentissage par renforcement.



Figure 10: Apprentissage par renforcement [34]

L'apprentissage par renforcement (RL) est un paradigme d'apprentissage automatique dans lequel un agent apprend à interagir avec un environnement pour atteindre un objectif en prenant des actions et en recevant des retours sous forme de récompenses ou de pénalités. L'objectif de l'agent est de maximiser la récompense cumulée qu'il reçoit au fil du temps.

En apprentissage par renforcement, l'agent apprend par essais et erreurs, explorant différentes actions et apprenant des conséquences de ces actions. Il vise à trouver une politique optimale, qui est une correspondance entre les états et les actions, qui maximise la récompense cumulative attendue.

D. Comparaison entre l'apprentissage supervisé et non supervisé

En tenant compte de ces paramètres ci-dessous, choisir l'apprentissage non supervisé peut être approprié lorsque l'objectif est d'explorer et de comprendre les structures et les modèles sous-jacents des données sans la nécessité de superviser explicitement le processus d'apprentissage. Dans la suite de notre document nous allons choisir **le clustering** comme méthode d'apprentissage non supervisé peut être bénéfique pour découvrir des structures intrinsèques des données, segmenter les données en groupes significatifs, détecter les anomalies, prétraiter les données et faciliter la visualisation et l'interprétation des données. Ces avantages en font une technique précieuse dans de nombreux domaines d'application de l'apprentissage automatique et de l'analyse de données.

Paramètres	Technique Apprentissage supervisé	Technique Apprentissage Non supervisé
Processus	Dans un modèle d'apprentissage supervisé, des variables d'entrée et de sortie seront données.	Dans le modèle d'apprentissage non supervisé, seules les données d'entrée seront données
Des données d'entrée	Les algorithmes sont formés à l'aide de données étiquetées	Les algorithmes sont utilisés contre des données qui ne sont pas étiquetées
Algorithmes utilisés	Prise en charge de la machine vectorielle, du réseau neuronal, de la régression linéaire et logistique, de la forêt aléatoire et des arbres de classification.	Les algorithmes non supervisés peuvent être divisés en différentes catégories : comme les algorithmes de cluster, les k-means, le clustering hiérarchique, etc.
Utilisation des données	Le modèle d'apprentissage supervisé utilise des données d'apprentissage pour apprendre un lien entre l'entrée et les sorties.	L'apprentissage non supervisé n'utilise pas de données de sortie.
Apprentissage en temps réel	La méthode d'apprentissage se déroule hors ligne.	La méthode d'apprentissage se déroule en temps réel.
Nombre de classe	Classifier les métadonnées peut être un véritable défi en apprentissage	Vous ne pouvez pas obtenir d'informations précises concernant
Inconvénient principal	Classifier les métadonnées peut être un véritable défi en apprentissage supervisé	Vous ne pouvez pas obtenir d'informations précises concernant le tri des données, et la sortie en tant que données utilisées dans l'apprentissage non supervisé est étiquetée et inconnue.

Tableau 4 : comparaison d'apprentissage supervisé et non [35]

III. Le clustering

Le clustering est une technique largement utilisée dans les réseaux IoT pour organiser les nœuds en groupes appelés clusters. Ces clusters permettent une gestion plus efficace des communications et des ressources dans le réseau.

Le clustering est une technique importante dans les réseaux IoT, offrant des avantages tels que la réduction de la surcharge du réseau, la conservation de l'énergie et une meilleure gestion des ressources. Il est largement utilisé dans une variété d'applications IoT et de surveillance pour améliorer l'efficacité et les performances des réseaux de capteurs. Voici une description de cette technique :

A. Caractéristique du clustering

Plusieurs caractéristiques de clustering sont improvisées pour catégoriser les différentes méthodes de clustering selon Arjunan et Sujatha[40]. Trois caractéristiques du clustering sont décrites dans cette sous-section :

- Propriétés des clusters
- Propriétés du CH
- Propriétés du processus de clustering

1. Propriétés des clusters

La taille du cluster, le nombre de clusters, la communication intra et inter-clusters sont utilisés pour définir les propriétés ou les spécifications des clusters.

- Nombre de clusters : le nombre de clusters désigne le nombre de clusters établis dans le RCSF, qui peut être prédéfini ou variable.
- Taille des clusters : dans les RCSFs, la taille des clusters peut être différente. Beaucoup d'entre eux peuvent être égaux ou inégaux[41]
- Communication intra-cluster : l'ensemble des communications entre les nœuds membres et le CH d'un cluster peuvent être direct ou multi-sauts.
- Communication inter- cluster : les communications entre clusters ou entre clusters et stations de base peuvent être directes ou multi-sauts.

2. Propriétés du CH

Les principales responsabilités du CH sont d'accumuler les données des membres du cluster, d'associer et d'agréger les données, puis d'envoyer les données collectées au SB. Il peut s'agir d'un système de communication directe ou multi-sauts.

- La mobilité : le cluster head peut être fixe ou mobile[42].
- Rôle : le cluster head (CH) a un rôle important. Il collecte les données de ses nœuds membres de son propre cluster, effectue l'agrégation des données des capteurs collectés et les envoie au prochain CH ou SB.

B. Propriétés du processus de clustering

Les propriétés du processus de clustering sont indiqués ci-dessous :

Méthodes de clustering : Il existe deux types de méthodes de clustering dans les RCSF. L'une est centralisée et une autre distribuée. Dans la méthode centralisée, une SB centrale autorisée où un super nœud gère l'ensemble du processus comme la sélection du CH et la formation du cluster, etc. Dans la méthode distribuée, il n'y a pas de contrôle central autorisé.

Nature : La méthode de clustering peut être de nature proactive, réactive ou hybride. Dans la méthode proactive, le CH transmet continuellement des données à la SB. Si les données sont détectées et disponibles et que le CH transmet les données à la SB, on parle de méthode réactive. Dans la méthode hybride, le CH transmet des données à la station de base à des intervalles de temps habituels.

Sélection des CH : le processus de sélection des CH est divisé en deux segments, probabiliste et déterministe. Dans le processus probabiliste, la sélection du CH peut être faite de manière aléatoire, sans tenir compte des considérations précédentes. Dans la méthode déterministe, le processus de sélection du CH peut être fait en tenant compte de divers paramètres.

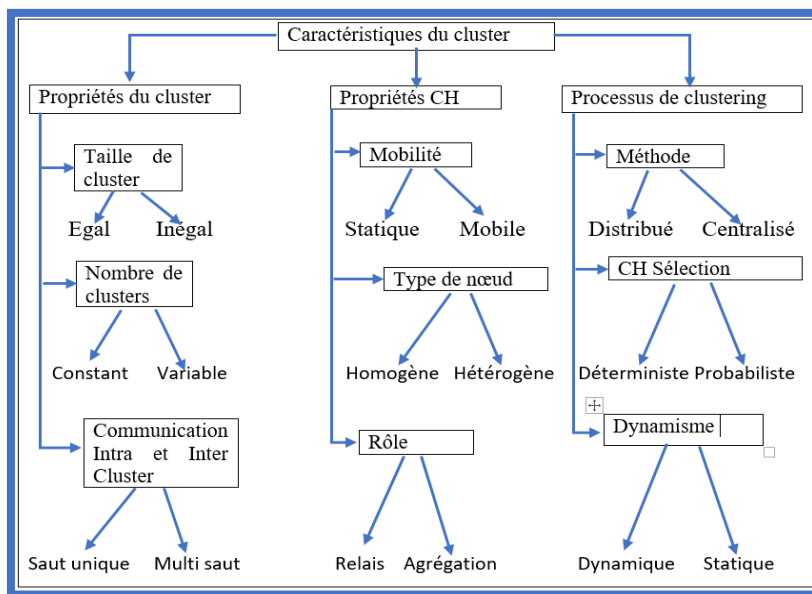


Figure 11: Les caractéristiques des algorithmes de clustering [42]

C. Les méthodes de clustering

Contrairement à la classification, le clustering est un processus d'apprentissage non supervisé où les classes ne sont pas connues à l'avance. Elle consiste à diviser les objets en groupes (clusters) en fonction des informations contenues dans les données décrivant les objets et leurs relations. Les objets appartenant au même cluster doivent être similaires les uns aux autres, mais doivent être aussi différents que possible des objets appartenant à d'autres clusters. Les techniques de clustering sont classées selon plusieurs aspects en méthodes de partitionnement, méthodes hiérarchiques, méthodes basées sur la densité, méthodes basées sur une grille et méthodes basées sur un modèle [43].

Dans les méthodes de partitionnement : plusieurs clusters sont formés à partir des données. Les algorithmes populaires de cette classe incluent k-means et PAM (Partitioning Around Medoids). L'algorithme k-means est l'une des techniques les plus couramment utilisées visant à segmenter une collection de n objets en k clusters (groupes). En agriculture, il est utilisé pour effectuer diverses tâches telles que la segmentation d'images pour la détection de maladies.

Les méthodes hiérarchiques : créent une hiérarchie de clusters appelée dendrogramme. Cette catégorie est divisée en méthodes de division ou cohésives. Les algorithmes de cette catégorie englobent CHAMELEON et BIRCH (Balanced Iterative Reducing and Clustering Using Hierarchies).

Les méthodes basées sur la densité : regroupent les objets selon une fonction cible de densité spécifique. Certains algorithmes de cette catégorie sont DBSCAN (Density-Based Spatial Clustering of Applications with Noise), DENCLUE (DENSITY CLUSTERING) et OPTICS (Ordering Points To Identify the Clustering Structure).

Les méthodes basées sur une grille : divisent l'espace en un nombre fini de cellules qui forment une structure de grille sur laquelle toutes les opérations de regroupement sont effectuées. Certains algorithmes de cette catégorie incluent STING (Statistical Information Grid approach) et CLIQUE (Clustering In Query).

Les méthodes basées sur des modèles : font référence à l'utilisation de modèles de mélanges finis pour effectuer des regroupements. COBWEB et CLASSIT sont des exemples d'algorithmes utilisés dans cette classe.

En ayant exploré les caractéristiques et les propriétés du clustering, ainsi que les méthodes de clustering disponibles, nous sommes maintenant prêts à examiner de plus près les algorithmes hiérarchiques et classiques et flous qui sont utilisés pour regrouper les objets de manière efficace et précise.

IV. Les algorithmes hiérarchiques

L'algorithmique hiérarchique, telle que l'algorithme LEACH (Low Energy Adaptive Clustering Hierarchy), ainsi que ses variantes, joue un rôle crucial dans la gestion efficace des réseaux de capteurs sans fil, notamment dans les réseaux de capteurs déployés pour l'Internet des objets (IoT). Voici une description générale de l'algorithme LEACH et de certaines de ses variantes.

A. LEACH

Le protocole de base est bien le protocole LEACH[44], son principe est la division du réseau en un ensemble de clusters selon la force reçue du signal et d'utiliser les CHs (chef de cluster) comme des routeurs pour passer les données à la station de base. Ce dernier prend en charge la gestion de son cluster. Il est élu périodiquement parmi les nœuds formant le cluster, en fonction de l'état de sa batterie.

Ce protocole permet ainsi la structuration du réseau de manière hiérarchique dans le but d'économiser l'énergie des capteurs.

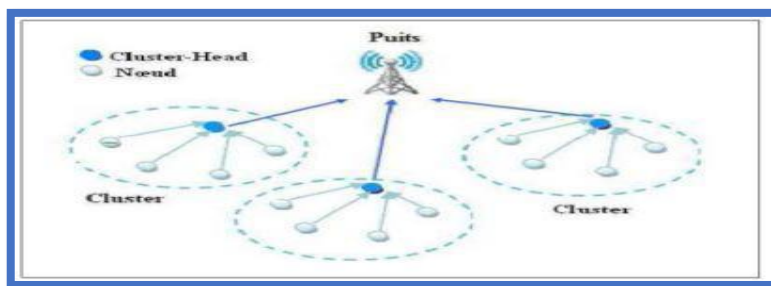


Figure 12: Routage hiérarchique basé sur le clustering [45]

Les dispositifs principaux du protocole LEACH sont :

- Coordination et contrôle localisés entre les nœuds : pour l'initialisation et le traitement de grappe
- Rotation randomisée de cluster : effectuée par "la station de base" ou "les têtes de cluster".
- Compression locale (agrégation) : Les nœuds CH compressent les données arrivant des nœuds appartenant à leur grappe respective, et envoi un paquet d'agrégation à la station de base afin de réduire la quantité d'information qui doit être transmise à la station de base.

Dans LEACH, le traitement est séparé dans des cycles de longueur constante, ou chaque cycle commence par une phase d'initialisation suivie d'une phase de transmission. La durée d'un cycle est déterminée.

1. La phase d'initialisation

Cette phase d'initialisation repose sur trois sous phases comme l'indique la figure 13 ci-dessous, la phase d'annonce, la phase d'organisation des groupes et la phase d'ordonnancement, qui permettent la formation des clusters et l'élection des CHs qui seront détaillées ci-dessous [45].

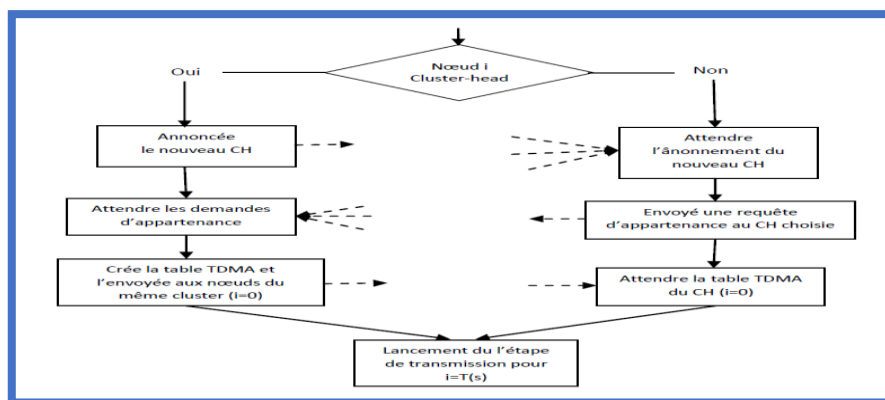


Figure 13: Opération de l'étape d'initialisation de LEACH [47]

2. La phase de transmission

Cette phase est plus longue que la phase précédente, et permet la collecte des données captées. En utilisant l'ordonnanceur TDMA, les membres émettent leurs données captées pendant leurs propres slots. Cela permet d'éteindre leurs interfaces de communication en dehors de leurs slots afin d'économiser leurs énergies. Ces données sont ensuite agrégées par les CH qui les fusionnent et les compressent, et envoient le résultat final au noeud puits. Après un certain temps prédéterminé, le réseau va passer à un nouveau round. Ce processus est répété jusqu'à ce que tous les nœuds de réseau seront élu CH. Une seule fois, tout au long des rounds précédents. Dans ce cas, le round est réinitialisé à 0.

3. Les avantages et limites de LEACH

Le protocole LEACH engendre beaucoup d'avantages en ce qu'il offre comme bonne manipulation de ressources de réseau en respectant plusieurs contraintes telles que la consommation d'énergie.

Bien que LEACH puisse augmenter la durée de vie du réseau, il présente certaines limitations. LEACH suppose que tous les nœuds ont la capacité de transmettre des données avec une grande puissance pour atteindre la station de base et que chaque noeud a une puissance de calcul lui permet de supporter différentes couches MAC. **Par conséquent, LEACH ne convient pas aux réseaux déployés dans de vastes régions.** En outre, LEACH choisit aléatoirement la liste des nœuds CH et il ne pose aucune contrainte sur leur distribution ainsi que sur leur niveau d'énergie. Ainsi, ces nœuds peuvent être concentrés dans un même endroit et par conséquent, il pourrait exister des nœuds du réseau isolés (n'ayant aucun noeud CH au voisinage). D'autre part, l'agrégation des données dans LEACH est centralisée et elle est exécutée périodiquement. Or, dans certains cas, la transmission périodique des données pourrait ne pas être nécessaire, ce qui épuise rapidement l'énergie limitée des nœuds capteurs.

4. Architecture de communication de LEACH

Tous les nœuds ordinaires transmettent leurs données à leur CH qui les agrègent et transmettent, à leur tour, à la BS selon une communication unicast (à un seul saut) [43].

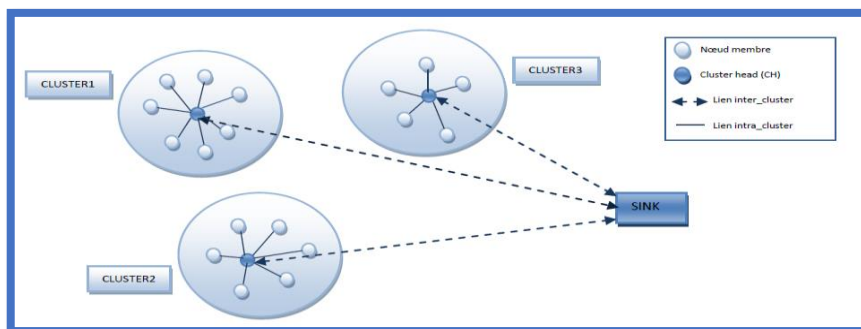


Figure 14: : Architecture de communication du protocole LEACH [45]

B. Les variantes du protocoles LEACH

En [février 2014] les auteurs [Manjusha M S et K E Kannammal] ont proposés une nouvelle amélioration du LEACH initial et différente des protocoles amélioré V-LEACH et V-LEACH AMÉLIORER. Cette amélioration touche principalement la technique de sélection du chef de cluster CH afin de réduire la consommation d'énergie.

Dans l'amélioration proposée, Ils ont aussi considéré l'architecture hiérarchique du réseau Des clusters Et les deux phases d'exécution du protocole La phase set-up et la phase steady-state.

La phase set-up va utiliser la nouvelle méthode de sélection du CH approprié Qui est responsable de la collection et l'agrégation des données depuis ces nœuds membres afin de les transmettre au BS.

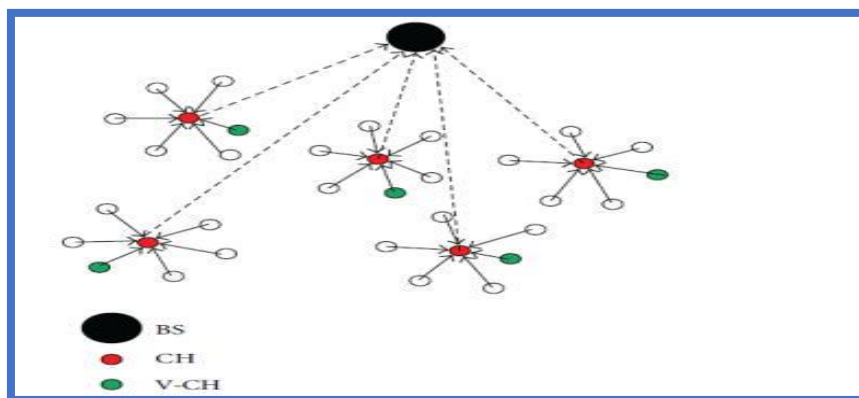


Figure 15: Processus de travail de LEACH amélioré [47]

Two-Vice LEACH protocole proposée par (Tebessi & Semchedine), est un améliorant au protocole V-LEACH, qui a fait ses preuves par rapport au protocole LEACH origine. Dans le protocole TV-LEACH ne se limite pas à utiliser un seul vice chef de cluster, mais plutôt deux. Ce protocole passe par plusieurs étapes pour l'installer et l'adapter comme la sélectionné les chefs de clusters et celle des vices chefs de clusters.

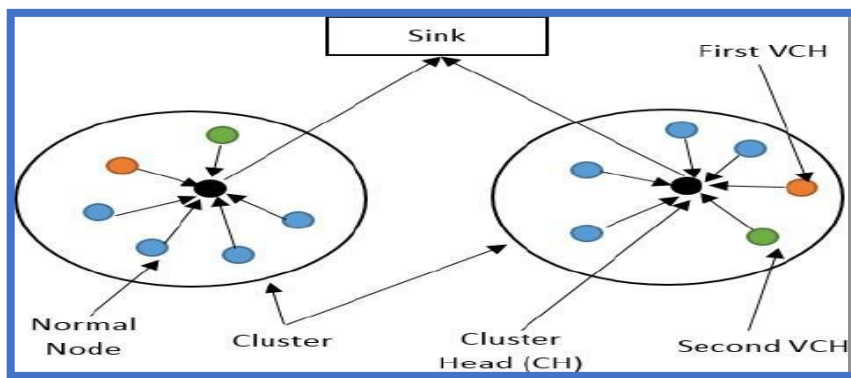


Figure 16: Architecture TV-LEACH [49]

1. Avantages et limites des protocoles dérivés de LEACH

Dans le tableau 5 ci-dessous les avantages et les limites des protocoles dérivés du protocole LEACH.

Protocole	Avantages	Limites
LEACH-C	Permet une diminution remarquable de la consommation énergétique	La version centralisée n'est pas adaptée aux réseaux de grande dimension
NN-LEACH	L'ensemble du réseau de capteurs peut être modélisé dans une perspective NN pour chaque noeud de capteur dans le réseau de capteurs, le NN décide le chef de cluster	La phase de sélection du CH consomme beaucoup de temps de calcul
Fire-LEACH	L'algorithme Firefly permet de fournir rapidement des solutions qui sont proches de la solution optimale	L'algorithme Firefly a d'inconvénient tel que se retrouver pris au piège dans plusieurs positions optimales
V-LEACH et TV-LEACH	Les données des nœuds de cluster atteindront toujours le SB ; pas besoin d'élire un	Frais supplémentaires pour l'élection de vice-CH et la formation des

	nouveau CH à chaque fois que le CH meurt. Cela prolongera la durée de vie du réseau	clusters
--	---	----------

Tableau 5 : Les algorithmes dérivés de LEACH

V. Les algorithmes classiques et flous

Les algorithmes classiques et flous de regroupement (clustering) tels que k-means, FOREL (Formiga, Répulsion, Éléphant, Letargia), et Fuzzy C-Means et ses variantes sont des techniques puissantes utilisées en apprentissage automatique et en analyse de données pour regrouper des ensembles de données en fonction de similarités intrinsèques. Chacun de ces algorithmes à ses propres caractéristiques et applications spécifiques dans le domaine de l'exploration de données.

A. L'algorithme k-means

L'algorithme k-means[46] est l'algorithme de clustering classique le plus connu et le plus utilisé, du fait de sa simplicité de mise en œuvre. Il partitionne les données d'une image en k clusters. Contrairement à d'autres méthodes dites hiérarchiques, qui créent une structure en « arbre de clusters » pour décrire les groupements, k-means ne crée qu'un seul niveau de clusters. L'algorithme renvoie une partition des données, dans laquelle les objets à l'intérieur de chaque cluster sont aussi proches que possible les uns des autres et aussi loin que possible des objets des autres clusters. Chaque cluster de la partition est défini par ses objets et son centroïde.

Soit $X = \{x_k / K = 1, \dots, n\}$ un sous-ensemble d'un espace n-dimensionnel. Une partition de X en C classes peut être représentée par des sous-groupes mutuellement disjoints C_1, C_c , tel que $C_1 \cup \dots \cup C_c = X$.

Chaque vecteur est assigné à une et seulement une classe qui lui est la plus proche.

La fonction objective correspondante s'exprime sous la forme :

$$J = \sum_{i=1}^c \sum_{k=1}^n U_{ik} d^2(x_k, v_i)$$

Où :

- v_i désigne le prototype (centroïde) de cette classe.
- d mesure la similarité entre le prototype et le vecteur de données.

- U_{ik} est un indicateur binaire $\in \{0, 1\}$ désigne l'appartenance ou non du vecteur x_k à la classe C_i ($U_{ik} = 1$ si $x_k \in C_i$ et $U_{ik} = 0$ sinon, $\forall i = 1 \dots C$ et $x_k \in X$)

L'indicateur d'appartenance U_{ik} s'exprime comme suit :

$$U(x_k, v_i) = \{1 \text{ si } d^2(x_k, v_i) = \min C(x_k, v_i) \text{ ou } 0 \text{ sinon}\}.$$

1. La sélection de cluster head (CH)

L'algorithme k-means qui forme les groupes d'objets en fonction des distances euclidiennes entre eux. Le schéma proposé pour la sélection du CH proposé comprend les trois étapes suivantes [47].

Étape 1 : clustering initial

L'algorithme k-means est exécuté pour la formation de clusters avec le WSN cible. Supposons que le WSN de n nœuds est divisé en k clusters. D'abord, k nœuds sur n sont choisis au hasard comme aléatoirement sélectionnés en tant que CHs. Chacun des nœuds restants décide de son CH le plus proche de lui en fonction de la distance euclidienne.

Étape 2 : Reclustering

Après avoir assigné chaque nœud du réseau à l'un des k clusters, le centroïde de chaque cluster est calculé. En supposant un espace bidimensionnel, le centroïde d'un cluster des nœuds est calculé comme suit.

$$\text{Centroïde}(X, Y) = \left(\frac{1}{s} \sum_{i=1}^s x_i, \frac{1}{s} \sum_{i=1}^s y_i \right)$$

Notez que le centroïde d'un cluster est un nœud virtuel se situant à la position centrale de l'amas. La figure 16 montre un exemple d'un cluster de 13 nœuds où le CH sélectionné aléatoirement au cours du tour initial est représenté par le nœud d'entouré. Observez sur la figure 10 que le CH initial n'est pas le plus proche du centroïde, et que le plus proche est donc le plus proche est sélectionné comme nouveau CH. Ici, il s'agit du nœud encerclé. Avec le nouveau CH dans chaque cluster, l'étape 2 est exécutée de manière récursive jusqu'à ce que le CH ne soit plus modifié.

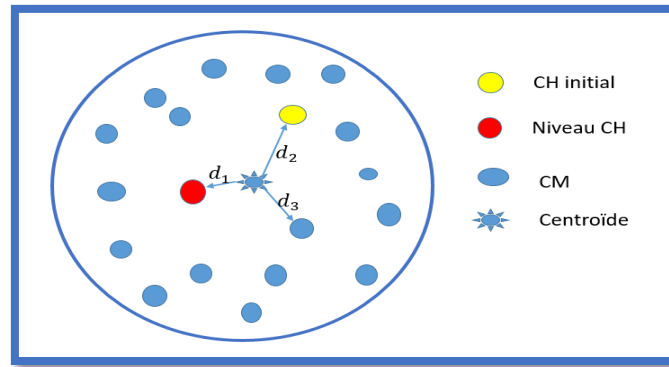


Figure 17: Formation d'un cluster

Étape 3 : Choix du CH

Une fois les clusters formés, un numéro d'identification est attribué à chaque nœud d'un cluster en fonction de la distance par rapport au centroïde, en attribuant un numéro plus petit à celui qui est le plus proche. La figure 18 montre l'ordre des nœuds de capteurs avec le numéro d'identification. Le numéro d'identification d'un nœud indique l'ordre dans lequel il sera choisi comme CH. Par conséquent, le numéro ID joue un rôle important dans la sélection d'un nœud comme CH.

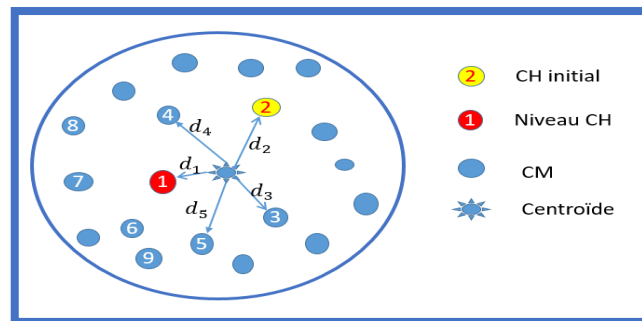


Figure 18: Commande des nœuds avec le numéro d'identification

NB : L'algorithme k-means, le choix des têtes de cluster est choisi aléatoirement dans les données d'entrées.

2. Les principes de l'algorithme k-means

L'optimisation se fait de manière itérative. Les étapes de cette classification pouvant être résumé comme suit :

- Initialiser l'ensemble des centroïdes avec des éléments, soit pris au hasard de l'ensemble de données, soit d'après des connaissances a priori.
- Assigner chaque point au centroïde le plus proche.

- Calculer les nouveaux centroïdes des classes construites. Les centroïdes étant les centres des classes, ils sont calculés par la moyenne des éléments de la classe.
- Générer une nouvelle partition de l'ensemble de données en assignant chaque point au centroïde le plus proche.
- S'il y a changement dans la partition, réitérer depuis l'étape 3.

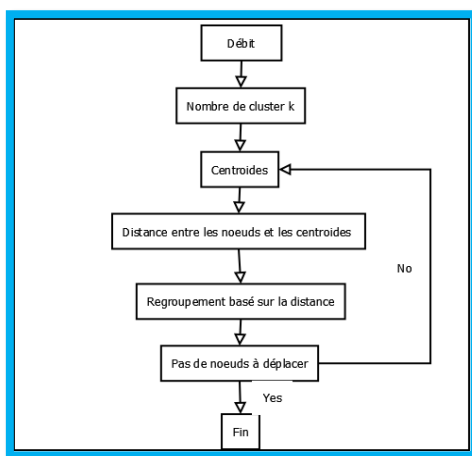


Figure 19: Organigramme de l'algorithme de k-means [52]

3. L'algorithme de k-means

Algorithme : k-means

Entrée

Ensemble de N données, noté par x
 Nombre de groupes souhaité, noté par k

Sortie

Une partition de K groupes {C1, C2, ... Ck}

Début

1) Initialisation aléatoire des centres Ck ;

Répéter

2) Affectation : générer une nouvelle partition en assignant chaque au groupe dont le centre est plus proche ;

$$X_i \in c_k \text{ si } \forall j |X_i - \mu_k| = \min |X_i - \mu_k|$$

Avec μ_k le centre de la classe c_k

3) Représentation : calculer les centres associés à la nouvelle partition ;

$$\mu_k = \frac{1}{N} \sum_{x \in c_k} X_i$$

Jusqu'à convergence de l'algorithme vers une partition stable ;

Fin.

L'équation utilisée

$$X_i \in c_k \text{ si } \forall j |X_i - \mu_k| = \min |X_i - \mu_j|$$

Avec μ_k le centre de la classe k

X_i : représente un objet dans l'ensemble des données et i .

c_k : représente le nombre de cluster et k le nombre de centre.

$|X_i - \mu_k|$: représente la distance entre le centre de la classe k et l'objet.

La formule utilisée pour calculer les centres associés à la nouvelle partition

$$\mu_k = \frac{1}{N} \sum_{x \in c_k} X_i$$

$\frac{1}{N} \sum_{x \in c_k} X_i$: représente la moyenne du cluster.

4. Les avantages et limites de k-means

❖ Les avantages de k-means :

Flexible : L'algorithme k-means s'adapte aux différentes modifications qui peuvent être apportées aux données.

Simple : L'algorithme k-means est facile à mettre en œuvre et identifie les ensembles de données inconnus à partir d'ensembles de données complexes. Les résultats seront affichés rapidement. **Convient aux gros ensembles de données** : Le protocole k-means convient à un grand nombre d'ensembles de données et est calculé plus rapidement que les plus petits.

Précision : L'analyse k-means garantit une précision de classification améliorée et fournit des informations sur un problème spécifique.

Clusters proches : Comparées aux algorithmes hiérarchiques, les k-means produisent des clusters plus serrés.

Efficace : k-means vous permet de diviser de grands ensembles de données. Son efficacité est représentée par les clusters qu'il forme.

Faible coût de calcul : Le coût de calcul de k-means est rapide et efficace.

❖ Les limites de k-means :

Manque de cohérence : La méthode de clustering k-means donne des résultats variables sur différentes exécutions d'un algorithme. Un choix aléatoire de modèles de clusters produit différents résultats, ce qui entraîne une incohérence.

Ensemble non optimal de clusters : k-means n'est pas permis de développer un groupe idéal et vous devez choisir les groupes avant d'obtenir des résultats efficaces.

Effet uniforme : Même lorsque les données d'entrée sont de tailles différentes, elles produisent un ensemble de tailles standard.

Traiter les données numériques : L'algorithme des moyennes k n'est implémenté que dans les données numériques.

Limitation des calculs : Lorsqu'il s'agit d'un grand ensemble de données, l'application d'une technique de dendrogramme fera planter l'ordinateur en raison de la charge de calcul et des limites de mémoire dynamique.

B. L'algorithme de clustering Forel

L'algorithme de Forel[48] a été introduit en 1967 et sa première application était liée au traitement statistique des données en paléontologie. FOREL appartient à une classe d'algorithmes dits de taxonomie. A ensemble de m objets peut être divisé en k taxons ($k < m$) de différentes manières. L'algorithme de Forel travaille avec les objets décrits par des caractéristiques quantitatives. Dans l'espace métrique, la distance euclidienne sert de caractéristique. Les objets inclus dans le même taxon sont placés dans une hyper sphère avec un certain centre C et un rayon R . En changeant le rayon, on peut obtenir un nombre différent de taxons.

Si le rayon R est fixé, l'algorithme est exécuté comme suit. Le centre C_1 est placé en un point quelconque de l'ensemble des objets. Ensuite, on détermine les points qui s'avèrent être à l'intérieur de la sphère. À cette fin, les distances d du point C_1 à tous les M sont calculées. Les points pour lesquels $d \leq R$ sont considérés comme internes. Le centre de gravité des points internes est calculé et le centre de la sphère est ensuite déplacé vers ce centre de gravité C_2 . Pour la nouvelle position, les points internes et leurs centres de gravité sont à nouveau trouvés. La procédure est répétée jusqu'à ce que les coordonnées du centre de gravité C_1 commencent à varier. On a l'impression que la sphère "nage" dans la direction de l'extrême local des points (voir Figure 20). Cette sphère est maintenant appelée taxon S_1 et ses points sont exclus de toute considération ultérieure.

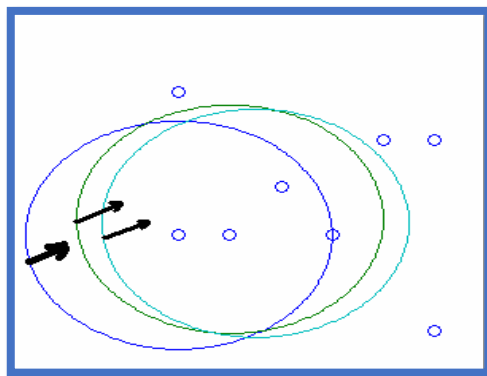


Figure 20: Déplacement de la sphère dans la direction du point extrême [20]

Ensuite, le centre d'une hyper sphère de même rayon est déplacé vers l'un des points restants et la procédure de révélation des taxons est répétée jusqu'à ce que tous les objets soient répartis entre les taxons. Il est évident que plus le nombre de taxons est grand, plus le rayon du taxon est petit. Le nombre de taxons souhaité par l'utilisateur peut être déterminé en ajustant correctement le rayon R .

C. L'algorithme de FCM

L'algorithme de clustering flou FCM attribue à chaque point de données une appartenance partielle à chaque cluster, plutôt qu'une appartenance binaire comme dans K-Means. La modélisation de l'imprécision s'effectue en considérant des frontières graduelles au lieu de frontières nettes entre les classes. L'incertitude s'exprime par le fait qu'un pixel possède aussi bien des attributs qui l'assignent à une classe qu'à une autre. La classification floue assigne donc, non pas à un pixel une étiquette relative à une classe unique, mais son degré d'appartenance à chacune des classes. Ces valeurs expriment l'appartenance incertaine d'un pixel à une région et sont appelées degrés d'appartenance [53]. Le degré d'appartenance se situe dans l'intervalle $[0, 1]$ et les classes obtenues ne sont pas forcément disjointes. Dans ce cas, les données X_j ne sont plus assignées à une classe unique, mais à plusieurs par l'intermédiaire de degrés d'appartenance U_{ij} du vecteur X_j à la classe i . Le but des algorithmes de classification est non seulement de calculer les centres de classe B mais aussi l'ensemble des degrés d'appartenance des vecteurs aux classes.

1. La méthode Fuzzy C_Means

Fuzzy C_Means (FCM)[50] est un algorithme de classification non-supervisée floue. Il est dérivé de l'algorithme des C_moyennes (C_Means) identique à l'algorithme K_Means, développé par Bezdek en 1981 à la suite des travaux de Dunn.

Dans la classification floue, un pixel n'est pas une étiquette relative à une classe unique, mais son degré d'appartenance à chacune des classes. Les degrés d'appartenance sont des valeurs qui expriment l'appartenance incertaine d'un pixel à une région, et sont compris entre 0 et 1, et les classes obtenues ne sont pas forcément disjointes.

Dans ce cas il est possible qu'une donnée x appartienne à deux ou plusieurs classes selon différents pourcentage « degrés d'appartenances ».

Pour faire l'algorithme FCM on a besoin de construire une matrice appropriée nommée μ dont les facteurs sont des nombre compris entre 0 et 1, et représentent le degré d'appartenance entre les centres et les données « pixels ».

$$\mu_{n \times m} = \begin{cases} 0.7 & 0.3 \\ 0.6 & 0.4 \\ 0.2 & 0.8 \end{cases}$$

La formule de la matrice des degrés d'appartenances :

$$\mu_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{d(x_i, c_j)}{d(x_i, c_k)} \right)^{2/(m-1)}}$$

Où

$d(x_i, c_j)$: C'est la distance entre le pixel x_i et le centre c_j , généralement on utilise la distance euclidienne.

m : est un nombre réel (> 1).

Cet algorithme nécessite la connaissance préalable du nombre de classes et les génère par un processus itératif en minimisant une fonction objective.

➤ La fonction objective :

$$J_{FCM}(C, U, X) = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \|x_i - c_j\|^2 \quad 1 \leq m \leq \infty$$

μ_{ij} : est le degré d'appartenance de x_i dans le j -ème classe.

x_i : est le i ème élément des données à classé.

c_j : est le centre d'une classe.

La formule pour calculer le centre :

$$C_j = \frac{\sum_{i=1}^N ((\mu_{ij})^m * x_i)}{\sum_{i=1}^N (\mu_{ij})^m}$$

2. Les étapes de l'algorithme FCM

Avant d'utiliser l'algorithme FCM, doit d'abord définir les paramètres suivants[55] :

C : Nombre de classe.

m : Le flou exposant (facteur flou).

$\varepsilon < 0.001$: Terme de pondération.

1) Initialiser les centres des C classes $c_j \mid j=1\dots C$ avec des éléments au hasard de l'ensemble de données. C : c'est le nombre des centres (classes).

2) Calculer la matrice des degrés d'appartenance $U_{ij} \mid i=1\dots N ; j=1\dots C$ on utilise la formule (9) et calculer la fonction d'objectif $J_{FCM}(c, U, X)$ par la formule (8). N : le nombre de pixels. X : les pixels. c : les centres. U : la matrice des degrés d'appartenance.

3) Mettre à jour les centres par la formule (10).

4) Mettre à jour la matrice U_{ij} et la fonction d'objectif.

Répéter les étapes 3 et 4 jusqu'à satisfaction du critère d'arrêt :

$$\|J_{FCM}(c, U, X)^{(t)} - J_{FCM}(c, U, X)^{(t+1)}\| < \varepsilon$$

3. Les avantages et les limites de Fuzzy C_Means

Les avantages :

- La méthode Fuzzy C_Means est une méthode de classification non supervisée, elle ne nécessite aucune information sur les données.
- Donne un meilleur résultat pour les ensembles de données de chevauchement et relativement mieux que k-means.
- Contrairement k-means où le point de données doit exclusivement appartenir à un centre de cluster ici le point de données est attribué membre à chaque centre de cluster à la suite dont le point de données peut appartenir à plus d'un centre de cluster.
- L'algorithme FCM est populaire et très performant.
- Il est efficace lorsque les objets sont bien séparés (surtout le font et les caractères).

Les limites :

- Pour le FCM, la décision d'appartenance d'un pixel à une classe n'est prise qu'à la fin de la convergence, contrairement aux k moyennes qui affectent un pixel à une classe à chaque itération.

- La sensibilité à la proposition initiale.
- La sensibilité au bruit.
- Long temps de calcul.

NB : Afin d'éviter ces limites donc améliorer les résultats de la classification, plusieurs modifications ont été apportées à l'algorithme et qui sont présentées comme des variantes du FCM.

4. Organigramme de FCM

L'organigramme de l'algorithme est la même sur tous ses améliorations. Mais les calculs des centres de classe, les matrices de degré d'appartenance et la fonction objective différent.

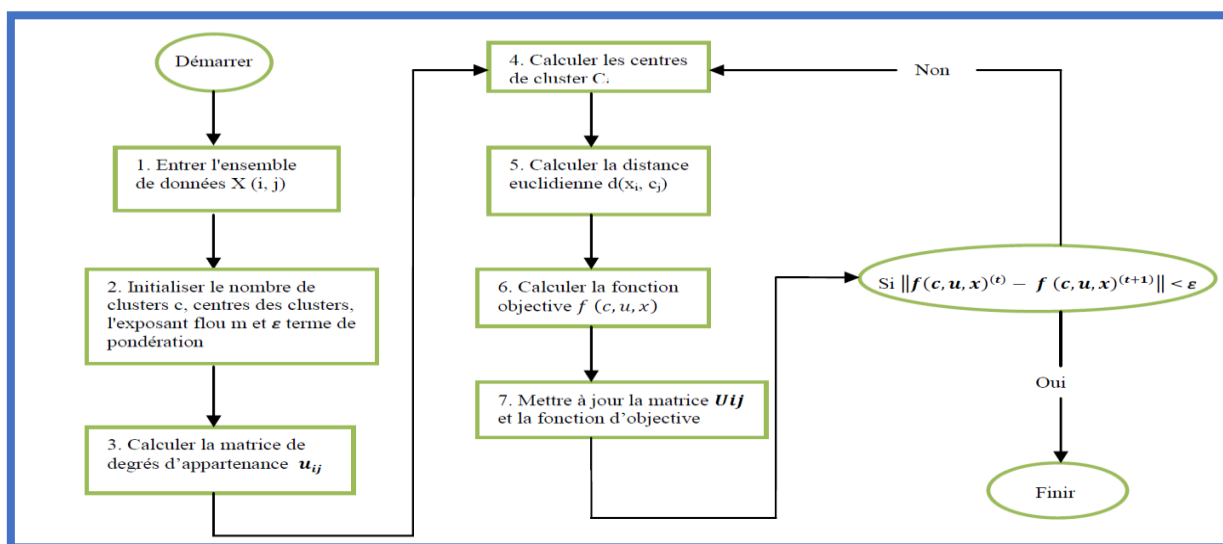


Figure 21: Organigramme de l'algorithme FCM [53]

Dans cette partie, nous avons présenté la méthode de segmentation floue FCM (Fuzzy C_Means), qui est née pour contester les méthodes classiques de segmentation, en associant à un pixel différentes classes selon des degrés d'appartenance au lieu de se limiter à une et une seule classe. Ainsi nous avons introduit quelques notions de base sur la logique floue, les fonctions d'appartenances et les opérations applicables sur les ensembles flous. Nous avons ensuite, présenté le principe de base de l'algorithme FCM, et avons étudié quelques-unes de ses variantes en se focalisant sur la mesure de distance et de la fonction objectif. L'étude est portée sur les modèles FCM- σ , PPCM et KFCM.

D. Les variantes de Fuzzy C_Means

De nombreuses variantes du modèle de base FCM ont été proposées dans la littérature, dans lesquelles de nouvelles mesures et variations sur les formules de distance et de la fonction objectif ont été introduites en vue d'aboutir à une segmentation efficace et suffisamment corrélée avec les entités qui composent l'image. Dans ce qui suit, nous présenterons quelques modèles ayant prouvé leur efficacité dans ce type d'approche [48, 34].

Les variantes selon la distance :

1. L'algorithme de Tsai et Lin FCM- σ

Tsai et Lin ont proposé une nouvelle mesure de distance qui prend en considération la variation des distances dans chaque classe pour améliorer l'efficacité de l'algorithme FCM. La distance est donnée par la formule suivante[51] :

$$d_{ij}^2 = \frac{\|x_i - c_j\|}{\sigma_j} ; i = 1..N, j = 1..C$$

σ_j : est la moyenne pondérée des distances dans la classe C, défini par :

$$\sigma_j = \left[\frac{\sum_{i=1}^C \mu_{ij}^m \|x_i - c_j\|^2}{\sum_{i=1}^C \mu_{ij}^m} \right]^{\frac{1}{2}}$$

La fonction objective ainsi est modifiée, elle est définie par :

$$J_{FCM}(c, U, X) = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m * \left(\frac{\|x_i - c_j\|^2}{\sigma_j} \right)$$

La matrice des degrés d'appartenance, est définie par :

$$\mu_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}}}, m \neq 1$$

Les centres de classes sont mis à jour comme dans l'algorithme FCM (se référer à l'équation

Ainsi, les étapes de la méthode $FCM\sigma$ se présentent comme suit :

- Initialiser les centres des classes x_i et initialiser la matrice des degrés d'appartenance μ_{ij} avec des valeurs entre [0,1] de manière aléatoire.
- Calculer la moyenne des distances σ_j par la formule (12).

- Calculer la matrice des degrés d'appartenance μ_{ij} par la formule (14) et la fonction objectif $J_{FCM}(c, U, X, \sigma)$ par la formule (13).
- Mettre à jour les centres des classes c_j par la formule (10) et la moyenne σ_j .
- Mettre à jour la matrice des degrés d'appartenance μ_{ij} par la formule (14) et la fonction objectif $J_{FCM}(c, U, X, \sigma)$ par la formule (13).
 - Répéter les étapes 4 et 5 jusqu'à satisfaction du critère d'arrêt :

$$\|J_{FCM}(c, U, X)^{(t)} - J_{FCM}(c, U, X)^{(t+1)}\| < \varepsilon$$
 t étant la t-ème itération.

2. L'algorithme de Kemel fuzzy C_Means KFCM

La fonction noyau est utilisée pour calculer la distance entre les pixels et les centres des classes, elle transforme les algorithmes linéaires en algorithmes non linéaires équivalents.

La fonction objectif est modifié comme suit [52] :

$$J_{FCM}(c, U, X) = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \|\varphi(x_i) - \varphi(c_j)\|^2$$

Ou $\|\varphi(x_i) - \varphi(c_j)\|^2$ est le carré de la distance entre $\varphi(x_i)$ et $\varphi(c_j)$

La distance entre les pixels et les centres des classes est donné par :

$$\begin{aligned} \varphi_{d_{ij}^2} &= \|\varphi(x_i) - \varphi(c_j)\|^2 \\ &= (\varphi(x_i) - \varphi(c_j))(\varphi(x_i) - \varphi(c_j)) \\ &= \varphi(x_i)\varphi(x_i) - 2\varphi(x_i)\varphi(c_j) + \varphi(c_j)\varphi(c_j) \\ &= K(x_i, x_i) - 2K(x_i, c_j) + K(c_j, c_j) \end{aligned}$$

En utilisant le noyau gaussien à base radiale :

$$K(x, y) = \exp\left(-\frac{\|x-y\|^2}{\delta^2}\right) \text{ Alors } K(x, x) = 1$$

δ Est présenté comme une dispersion, nous utilisons la variance d'échantillon pour estimer δ^2 .

$$\text{Avec : } \delta^2 = \frac{\sum_{i=1}^N \|x_i - \bar{x}_i\|^2}{N} \quad \text{et} \quad \bar{x}_i = \frac{\sum_{i=1}^N x_i}{N-1}$$

Et la fonction objective devient comme :

$$J_{KFCM}(c, U, X) = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \|1 - K(x_i, c_j)\|^2$$

La matrice des degrés d'appartenance est définie comme :

$$\mu_{ij} = \sum_{k=1}^c \frac{1}{\sum_{k=1}^c \left(\frac{\|1 - K(x_i, c_j)\|^2}{\|1 - K(x_i, c_k)\|^2} \right)^{1/m-1}}$$

Les centres des classes C_j sont définis comme :

$$C_j = \frac{\sum_{i=1}^N ((\mu_{ij})^m * x_i)}{\sum_{i=1}^N (\mu_{ij})^m}$$

3. L'algorithme de C moyennes Possibiliste (PCM)

Krishnopuram et Keller ont proposé une approche de possibiliste des C_moyennes appelé Possibilistic C-Means en anglais (PCM) qui introduit la notion de possibilité très voisine du flou. Le degré d'appartenance d'une observation à une classe est considéré comme une possibilité d'appartenir à une classe. Contrairement à la méthode FCM, la somme des degrés d'appartenance d'un pixel à toutes les classes n'est pas égale à 1.

La fonction objectif est modifiée comme suit [43] :

$$J_{FCM}(c, U, X, f) = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m * d^2(x_i, c_j) + \sum_{j=1}^C f_j \sum_{i=1}^N (1 - \mu_{ij})^m$$

$$\sum_{i=1}^N (\mu_{ij})^m > 0 \forall j \in \{1, \dots, C\}$$

f_j : Nombre positif à déterminer.

La première partie de l'équation () est identique à celle de FCM.

La seconde partie force μ_{ij} à la plus grande valeur possible.

On obtient l'équation de mise à jour des degrés d'appartenance à partir de la formule

Suivant :

$$\mu_{ij} = \frac{1}{1 + \left[\left(\frac{d^2(x_i, c_j)}{f_j} \right)^{1/m-1} \right]}$$

Sachant que le paramètre f_j est estimé à chaque itération, en le calculant comme suit :

$$f_j = \frac{\sum_{i=1}^N (\mu_{ij})^m * d_{ij}^2(x_i, c_j)}{\sum_{i=1}^N \mu_{ij}^m}$$

Les centres de classes sont mis à jour comme dans l'algorithme FCM.

VI. Comparaison des propriétés des algorithmes

Dans cette partie, nous proposons d'effectuer une comparaison des différents algorithmes de clustering décrits précédemment selon plusieurs métriques : le type (1-hétérogène ou 2-homogène), métriques du nombre de CH (déterminé ou indéterminé), la connectivité intra-cluster et inter-cluster, méthode de clustering (distribuée ou centralisée), la prise en charge de la mobilité des nœuds et l'adaptation de l'algorithme (petits ou grands réseaux).

Les algorithmes traditionnels centralisés d'intra-clustering nécessitent souvent la connaissance de l'ensemble des informations du réseau. La perte d'un nœud critique ou une erreur de communication peut éventuellement provoquer un échec grave du clustering.

A. Intra-clustering

La communication entre le nœud CH et les autres nœuds membres du cluster peut se faire, soit en un seul saut soit, en plusieurs sauts.[54]

- Communication direct (en un saut) : les paquets de données sont envoyés directement au CH. Ce type de communication engendre une consommation importante d'énergie, si la distance entre le CH et les nœuds est grande.
- Communication indirect (en plusieurs sauts) : Dans ce cas, chaque membre du cluster envoie ses données au plus proche membre de son cluster jusqu'à l'aboutissement au CH. Ce type de communication engendre une consommation minimal d'énergie.

L'intra-clustering à un saut constitue une solution pratique pour structurer le réseau. Cependant, pour les grands réseaux, de nombreux clusters sont formés. Par conséquent, des dispositifs isolés peuvent apparaître et la connectivité peut être difficile à assurer. L'intra-clustering multi saut (ou k-sauts) atténue le problème en produisant une structure cohérente dans des environnements dynamiques et à grande échelle. Dans cette approche, les membres du cluster envoient leurs données aux têtes de cluster par le biais d'une communication multi-sauts et le CH agrège les données directement à la BS. L'objectif est de réduire la consommation d'énergie intra-cluster en utilisant des communication multi-sauts [55].

B. Inter-clustering

Les CHs peuvent communiquer directement avec la station de base, ou en plusieurs sauts via des nœuds appelés des nœuds Gateway. Il peut s'agir des nœuds CHs ou des nœuds membres du cluster. L'utilisation de la communication multi-sauts réduit la consommation d'énergie et rend le réseau extensible [56].

Contrairement à l'algorithme LEACH, la procédure Fuzzy C-means (FCM) réalise un routage multi-saut entre les CHs et la station de base. Le choix de la bonne route dépend des poids des liens qui dépendent à leur tour des distances entre les nœuds, de nombre de messages traversés et de l'énergie consommée.

Pour LEACH, le nombre moyen de paquets envoyés est faible, il représente en fait le nombre de CHs de la structure construite divisé par le nombre total de nœuds. En effet, chaque CHs envoie ses données directement vers la station de base [57].

Pour la procédure k-means, la communication des CHs et la station de base se fait à saut unique qui est de même pour Forel.

Cependant l'inter-clustering est la communication entre CHs et entre les CHs et la station de base. Une communication entre les CHs pour transmettre le message à la station de base est dite routage à multi saut et elle est à saut unique lorsque les CHs envoient le message directement à la station de base.

C. Equilibrage des charges

Dans les réseaux denses et à faible mobilité, la priorité n'est pas donnée à la stabilité de la topologie de clustering mais à l'équilibrage de la taille des clusters. La taille du cluster présente un compromis entre les performances de communication inter-cluster et intra-cluster. De plus, les grands clusters ont un impact considérable sur la consommation d'énergie des clusterheads, c'est pourquoi de nombreuses propositions tentent d'optimiser la sélection des clusterheads afin d'obtenir un ensemble de clusters de taille optimale [58]. Chaque capteur appartient à un et un seul Gateway et ce Gateway est son seul moyen pour communiquer avec la station de base. Afin d'assurer l'équilibrage de la charge, les super-nœuds doivent avoir toutes les informations concernant le réseau et de décider la taille optimale pour un cluster ; ceci nécessite un temps de collecte d'informations non négligeable. En plus, l'algorithme nécessite que chaque nœud soit équipé d'un système de localisation tel qu'un GPS ; ce qui s'avère coûteux financièrement.

D. Considération énergétique

La consommation d'énergie représente un élément clé pour la conception des réseaux de capteurs sans fil[54] et celui du technologie LoRaWAN. Plusieurs facteurs peuvent être une source de surconsommation d'énergie : la mobilité, la position d'un noeud dans le réseau (relais ou gateway), les retransmissions, etc... Le poids des nœuds considérés au cours de la phase d'élection des cluster-heads influt sur la quantité d'énergie consommée. La topologie du réseau peut être aussi une cause majeure sur la consommation d'énergie. La transmission en 1-saut peut endommager les batteries si la distance est assez élevée [59]. La transmission en k-sauts résout le problème mais affaiblit l'énergie des nœuds intermédiaires. On peut comparer l'énergie consommée par un cluster-head en effectuant la procédure d'agrégation des messages reçus avec celle consommée sans agrégation. Ils ont par la suite montré que lorsque l'énergie considérée pour l'agrégation est inférieure à une valeur limite ($1\mu\text{J/bit/signal}$), alors, la transmission avec agrégation nécessite une énergie plus faible que celle sans agrégation.[60] L'énergie consommée par un clusterhead lors de la transmission des informations vers la station de base augmente proportionnellement mais faiblement avec la valeur de la portée ; ceci s'explique par le fait que l'augmentation de la portée d'un noeud aboutit à une augmentation de la puissance d'émission et donc à l'augmentation de l'énergie consommée.

Nous vous présentons un tableau de comparaison des propriétés des algorithmes de clustering.

Algorithme	Nombre de CH	Topologie	Intra cluster	Inter cluster	Equilibrage des charges	Considération énergétique	Type	Adaptation aux réseaux LoRa
k-means	Déterminé	Centralisé	Saut unique	Saut unique	Oui	Non	2	Oui
Fuzzy c-means	Indéterminé	Centralisé	Saut unique	Multi-sauts	Oui	Oui	2	Oui
Forel	Déterminé	Centralisé	Multi-sauts	Saut unique	Non	Oui	2	Oui
KM-LEACH	Déterminé	Distribué	Saut unique	Saut unique	Oui	Oui	2	Oui
LEACH-FCM	Indéterminé	Distribué	Saut unique	Saut unique	Oui	Non	2	Oui

LEACH	Indéterminé	Distribué	Multi-sauts	Saut unique	Non	Oui	2	Non
--------------	--------------------	-----------	-------------	-------------	-----	-----	---	-----

Tableau 6 : Comparaison des algorithmes de clustering

D'après le tableau 6 ci-dessus nous constatons que certains algorithmes de clustering comme k-means, FCM et Forel sont adaptés dans les réseaux LoRa. Elles sont utilisées pour la transmission de données sur de grandes distances avec une faible consommation d'énergie.

Cependant si on utilise LoRa dans un réseau de capteurs et qu'on envisage d'appliquer k-means pour la collecte et l'analyse de données, ainsi que LEACH pour la gestion de l'énergie, cela pourrait être une approche valable, mais la combinaison exacte dépendra des besoins spécifiques de votre application.

De plus l'adaptabilité de la combinaison LEACH-FCM peut dépendre des exigences spécifiques de votre application et de la nature de votre ensemble de données. Avant de mettre en œuvre cette combinaison, il est recommandé de réaliser une évaluation approfondie des caractéristiques de votre réseau, de définir clairement les objectifs de regroupement et de prendre en compte les contraintes liées à LoRa.

D'après la littérature nous constatons que k-means est relativement simple à comprendre et à mettre en œuvre. Il est plus efficace pour des ensembles de données de taille modérée avec des clusters bien définis. Mais aussi il fonctionne bien même dans des espaces de données de grande dimension. Son temps de calcul peut augmenter linéairement avec le nombre de points de données, le nombre de clusters, et le nombre d'itérations nécessaires pour la convergence.

Conclusion

Les techniques de clustering dans les réseaux IoT ont une importance cruciale pour optimiser l'utilisation des ressources limitées dans ces réseaux. En utilisant des techniques de clustering, les nœuds du réseau peuvent être regroupés en clusters de manière efficace, permettant ainsi une gestion efficace des ressources telles que la bande passante, l'énergie et le stockage.

L'utilisation de techniques de clustering telles que LEACH, K-means, FCM et d'autres peut aider à prolonger la durée de vie des batteries des nœuds du réseau, à réduire les coûts de communication et à améliorer l'efficacité des opérations du réseau. Ces techniques sont utilisées

dans une variété d'applications IoT, allant de l'agriculture à la surveillance environnementale en passant par l'industrie, la santé et bien d'autres domaines.

En somme, les techniques de clustering jouent un rôle clé dans l'optimisation de la performance et de l'efficacité des réseaux IoT, en permettant une gestion efficace des ressources et une utilisation optimale des données collectées. Ainsi les chercheurs ont montré que LEACH ne convient pas aux réseaux déployés dans de vastes régions or LoRa est un réseau conçu pour de grande zone géographique.

Il est important de noter que le choix de l'algorithme dépend fortement du contexte particulier de l'application, des caractéristiques des données, et des objectifs spécifiques de l'analyse. Ainsi, le meilleur algorithme peut varier en fonction de la nature des données, de la complexité des clusters, des contraintes de ressources, et d'autres considérations spécifiques à chaque cas d'utilisation.

À la suite de notre exploration des techniques de clustering dans les réseaux IoT, nous nous plongeons désormais dans l'application spécifique de l'algorithme de clustering k-means dans le contexte des réseaux LoRa, offrant ainsi une perspective approfondie sur la gestion efficace des données générées par ces dispositifs IoT à longue portée.

Chapitre 3 : L'algorithme de clustering k-means Appliqué dans LoRa

Introduction

Le réseau LoRa (Long Range) est une technologie de communication sans fil basée sur la transmission de données à bas débit et à longue portée, ce qui en fait une solution idéale pour les applications IoT telles que la surveillance environnementale, la gestion de l'agriculture, l'industrie, etc. Pour maximiser l'utilisation des ressources limitées dans les réseaux LoRa, il est nécessaire d'utiliser des techniques de clustering pour regrouper les nœuds en clusters efficaces.

Dans ce contexte, les algorithmes de clustering appliqués dans LoRa jouent un rôle crucial pour permettre une gestion efficace des ressources telles que la bande passante, l'énergie et le stockage. Ces algorithmes sont conçus pour regrouper les nœuds LoRa en clusters basés sur des caractéristiques communes, telles que la distance, la puissance du signal, le niveau de batterie, etc. Dans cette optique, ce chapitre se concentre sur l'analyse des différents algorithmes de clustering appliqués dans les réseaux LoRa, en soulignant leurs avantages et leurs limites pour la gestion des ressources et l'optimisation des performances du réseau.

I. Utilisation de k-means dans LoRa

L'utilisation de l'algorithme de clustering k-means dans les réseaux LoRa soulève certaines problématiques spécifiques. Voici quelques-unes des problématiques courantes liées à l'utilisation de l'algorithme k-means dans les réseaux LoRa.

La plupart des contrôles de puissance de transmission sont concernés l'augmentation de la capacité, et pas seulement la diminution de la consommation d'énergie. Le contrôle du débit de transmission est souvent préoccupé par la maximisation du débit [59]. Par rapport à LoRa, les débits de paquets WiFi sont nettement plus élevés, et les algorithmes de contrôle ADR s'exécutent à un rythme beaucoup plus élevé qu'en LoRa. Par exemple, l'algorithme de contrôle du taux de transmission le plus couramment le plus utilisé, évalue ses liens toutes les 100 ms.[60] Dans cette partie nous présentons les articles existants dans la littérature qui ont travaillé dans cette approche.

Agrégation de paquets : l'article de [62] propose une approche systématique pour traiter les données réseau recueillies d'une plateforme sans fil IoT à grande échelle basée sur LoRaWAN (Long Range Wide Area Network). Cette étude peut être utilisée pour profiler les dispositifs IoT, afin de les regrouper en fonction de leurs caractéristiques, ainsi que pour détecter les

anomalies du réseau. Plus précisément, ils utilisent l'algorithme k-means pour regrouper les paquets LoRaWAN en fonction de leur comportement radio et réseau. Ils ont testé cette approche sur un réseau LoRaWAN réel où l'ensemble du trafic capturé est stocké dans une base de données propriétaire. Tout à fait important est le fait que LoRaWAN capture, via l'interface sans fil, des paquets de plusieurs opérateurs. En effet, cette analyse a été effectuée sur 997 183 paquets avec 2169 dispositifs impliqués et seul un sous-ensemble d'entre eux était connu de l'opérateur considéré, ce qui signifie qu'un opérateur ne peut pas contrôler le trafic de l'opérateur considéré l'ensemble du comportement du système mais doit au contraire l'observer [63]. Ils ont pu analyser le contenu des clusters, révélant des résultats à la fois conformes au comportement actuel du réseau et des alertes sur des dispositifs défectueux, soulignant la fiabilité de l'approche proposée.

Allocation du facteur d'étalement (SF) : l'article de [65] contribue au problème d'allocation du facteur d'étalement LoRa (SF) en proposant un algorithme basé sur le clustering k-means. Nous évaluons la performance du réseau en tenant compte des probabilités d'interruption d'un réseau LoRaWAN de classe A à grande échelle, en mode non confirmé, sans retransmissions.

Certaines méthodes telles que EXP-SF et EXP-AT, évoquées dans la référence [66], sont conçues pour améliorer le débit global des données. En revanche, EXPLoRa-TS, décrit dans la référence [67], propose une répartition équilibrée de la charge de trafic entre les facteurs d'étalement tout en tenant compte du fait que chaque appareil final transmet une quantité de données variable en fonction de son profil d'application.

D'autres recherches se concentrent sur l'application de techniques d'apprentissage automatique, comme mentionné dans les références [67], [68], et [69]. Par exemple, dans les articles [68] et [69], on a appliqué une méthode d'apprentissage non supervisée de type k-means pour identifier des clusters appropriés en vue de l'allocation des facteurs d'étalement. Dans d'autres publications, on explore l'utilisation de l'apprentissage par renforcement, où une passerelle agit comme un agent pour contrôler l'environnement (les appareils finaux) en envoyant des actions (configurations). En outre, des réseaux neuronaux convolutifs ont été utilisés dans [67] pour résoudre des problèmes d'optimisation.

Contraintes de communication : Les réseaux LoRa ont des contraintes de bande passante et de consommation d'énergie. L'algorithme k-means implique souvent une communication fréquente entre les dispositifs pour échanger des informations sur les mesures de signal et les

clusters.[66] Cela peut entraîner une surcharge de communication dans les réseaux LoRa, ce qui peut être problématique en termes de bande passante limitée et de consommation d'énergie.

Variabilité des signaux : Les signaux LoRa peuvent être sujets à des variations, des interférences et des atténuations dans l'environnement. Cela peut rendre les mesures de signal moins fiables, ce qui peut affecter la précision du clustering [67]. L'algorithme k-means repose souvent sur la distance euclidienne pour calculer les similarités entre les points de données, ce qui peut être affecté par la variabilité des signaux LoRa.

II. Les limites des solutions existantes

Les solutions basées sur l'algorithme de clustering k-means dans les réseaux LoRa présentent également certaines limites spécifiques. Voici quelques-unes des principales limites à prendre en compte dans certains articles :

Contraintes de déploiement : Dans l'article [65], les auteurs ont présenté une zone de couverture. En limitant la zone à un rayon de 3 km, à l'intérieur de ce cercle, ils créent des bandes circulaires définies par des limites extérieures et intérieures. Chaque bande se voit attribuer une valeur de facteur d'étalement. Ainsi, tous les nœuds situés en dehors de la bande sont affectés par un SF7. Cependant, cette proposition pourrait rencontrer des difficultés dans une zone d'un rayon de 4 à 5 km.

Contraintes de bande passante : Les réseaux LoRa utilisent différentes valeurs du facteur d'étalement pour gérer la portée de transmission et la consommation d'énergie. Cependant, chaque facteur d'étalement utilisé nécessite une certaine quantité de bande passante. Dans les réseaux LoRa à grande échelle avec de nombreux dispositifs, l'allocation des facteurs d'étalement devient un défi en raison des contraintes de bande passante limitée.[65] Il peut être difficile de trouver un équilibre entre la portée de transmission, la qualité de la communication et l'utilisation efficace de la bande passante disponible.

Interférences et collisions : L'allocation inadéquate des facteurs d'étalement peut entraîner des interférences et des collisions entre les transmissions des dispositifs LoRa. Lorsque plusieurs dispositifs utilisent le même facteurs d'étalement dans une zone géographique proche, cela peut causer des collisions et réduire la fiabilité des communications [65]. L'allocation efficace des facteurs d'étalement doit tenir compte de la densité des dispositifs et des zones de recouvrement pour minimiser les interférences et maximiser les performances du réseau.

Capacité limitée : Lorsque plusieurs clusters utilisent le même facteur d'étalement, cela peut limiter la capacité du réseau à gérer simultanément un grand nombre de nœuds ou de dispositifs connectés. Les ressources du réseau, comme la bande passante disponible, peuvent devenir insuffisantes pour prendre en charge une demande élevée, ce qui peut entraîner des problèmes de latence et de débit réduit.

Coût énergétique : L'utilisation des facteurs d'étalement plus élevés dans les dispositifs LoRa peut augmenter la consommation d'énergie lors des transmissions [65]. L'allocation des facteurs d'étalement doit trouver un équilibre entre la portée de transmission, la consommation d'énergie et la qualité de la communication. Une mauvaise allocation des facteurs d'étalement peut entraîner une utilisation inefficace de l'énergie et réduire la durée de vie des batteries des dispositifs.

D'autres études se concentrent principalement sur l'évaluation des performances des réseaux LoRaWAN, comme en témoignent les références [68] et [69], sans apporter de nouvelles méthodes ou solutions pour leur amélioration. La plupart de ces documents examinent des aspects tels que la densification du réseau, la couverture de la communication et l'efficacité énergétique dans le but de déterminer la durée de vie des batteries, mais ils ne se penchent pas sur les questions fondamentales de LoRa. En réalité,

Ils omettent de prendre en compte, dans leurs modèles mathématiques ou simulations, Certains aspects physiques cruciaux propres à LoRa, tels que l'effet de capture (interférences au sein d'un même facteur d'étalement) et l'orthogonalité imparfaite des facteurs d'étalement (interférences entre différents facteurs d'étalement) [66], [68], [70]. Ces phénomènes ont un impact direct sur les performances du réseau.

Ils se basent essentiellement sur le modèle de perte de propagation Log-Distance [68], [71], [72], en supposant une perte de propagation pour une distance de référence d_0 de 40 mètres équivalente à 127.41 dB. Cette supposition est restrictive, limitant ainsi le déploiement aux environs d'une passerelle sur de courtes distances, ce qui va à l'encontre de l'objectif central des technologies LPWAN.

Ils sous-estiment l'influence d'autres facteurs cruciaux, tels qu'une charge de trafic élevée ou la dimension importante des réseaux (nombre de dispositifs dans les réseaux LoRaWAN à fort trafic [72]). Après avoir pris du recul pour examiner en profondeur les solutions existantes, nous avons jugé nécessaire de réexaminer la question de l'allocation des

paramètres radio, en particulier pour élaborer de nouvelles stratégies adaptées aux systèmes LoRaWAN soumis à une forte charge et aux réseaux de grande envergure.

III. Problématiques

Les réseaux LoRaWAN utilisent différentes valeurs de SF pour la modulation des signaux. Cependant, dans une zone géographique donnée, l'utilisation de SF identiques par plusieurs dispositifs peut entraîner des interférences, réduisant ainsi les performances du réseau. L'émergence rapide des réseaux IoT (Internet des Objets) a conduit à une adoption croissante de technologies de communication à longue portée, parmi lesquelles LoRa (Long Range) occupe une place prépondérante. Cependant, la fiabilité de la transmission des données dans les réseaux LoRa est confrontée à des défis importants, parmi lesquels les interférences entre les facteurs d'étalement représentent une problématique cruciale. Les facteurs d'étalement, déterminants pour la propagation du signal, deviennent des points sensibles susceptibles d'affecter la stabilité des communications et la capacité du réseau à gérer efficacement un grand nombre de dispositifs. Les dispositifs LoRa utilisent des facteurs d'étalement spécifiques pour moduler leurs signaux, permettant ainsi une coexistence harmonieuse des multiples appareils au sein du même réseau. Cependant, dans des environnements denses où plusieurs réseaux LoRa cohabitent ou lorsque d'autres technologies de communication partagent la même bande de fréquences, les interférences entre les facteurs d'étalement peuvent entraîner des perturbations significatives. Dans ce cas le problème de l'interférence de l'article [65] nous donc d'une proposition pour l'amélioration de l'algorithme proposé de cet article. L'algorithme proposé par les auteurs est un algorithme d'allocation de SF dans chaque cluster dans différentes bandes.

A. Algorithme de l'article de référence

Dans cet article, les auteurs ont proposé un algorithme d'allocation des SF, c'est-à-dire un algorithme permettant de définir l'étendue de chaque anneau de SF. Leur solution fait appel à l'algorithme d'apprentissage automatique k-means, utilisé dans le processus de quantification des vecteurs dans l'exploration des données par regroupement. Il s'agit d'une approche non déterministe, numérique et itérative. L'objectif principal de l'algorithme k-means est de trouver la fonction de coût minimum, définie comme la distance entre chaque point de l'ensemble de données et son centroïde le plus proche. L'algorithme proposé est la suivante :

Algorithme 1 Allocation de SF dans k-means[65]

Entrer ED := N nœuds déployés uniformément

Sortir L := $\{l_0, l_1, \dots, l_6\}$

l_6 := R

Pour i allant de 5 à 1 faire ▷ Pour chaque anneau SF en commençant par l'anneau le plus extérieur

K := `getkfromseries(i)` ▷ Définir le nombre de centroïdes k pour cette itération

C := `kmeans(ED,K)` ▷ Calculer le centre de gravité

B := `limite(C)` ▷ Calculer la limite de C

I := $\{x \in ED \mid x \in \text{conv}B\}$ ▷ Sélectionner les nœuds qui se trouvent à l'intérieur de la limite B

$l_i := \frac{\max(|I_x|) + \max(|I_y|)}{2}$ ▷ Calculer la nouvelle limite d'anneau SF

$SF_{i+7} = \{x \in ED \mid x \notin \text{Ball}[(0,0), I_i]\}$ ▷ Allouer SF_{i+7} aux nœuds en dehors du cercle de rayon I_i

ED := $\{x \in ED \mid x \in \text{Ball}[(0,0), I_i]\}$ ▷ Supprimer les nœuds en dehors du cercle de rayon I_i

$l_0 := 0$

$SF_7 := ED$ ▷ Allouer SF_7 aux nœuds restant

Retour L

Toutes les itérations de l'algorithme proposé pour un exemple de réseau. Le rayon de la zone circulaire du réseau est $R = 3\text{km}$, et par conséquent, la limite extérieure de SF12 est $l_6 = R = 3\text{km}$. La première itération de l'algorithme définit la limite intérieure du SF12, l_5 . Après avoir exclu de l'EDk les dispositifs situés à l'intérieur de l'anneau SF12, l'algorithme lance une nouvelle itération et définit l_4 , c'est-à-dire la limite intérieure de l'anneau SF11. Les itérations se poursuivent jusqu'à ce que l_1 soit défini et que la géométrie complète du réseau soit obtenue. Dans la suite nous allons détailler les problématiques de l'article de références.

Affectation SF dans une même bande : Pour chaque SF dans la zone de couverture, les auteurs ont estimé le nombre de cluster optimal cela montre que les clusters qui sont dans une même bande ont le même SF. Cela a des conséquences du fait que sur une bande de plusieurs têtes de cluster la passerelle reçoit plusieurs messages qui aura pour cause des pertes de paquets au niveau de la passerelle qui peut entraîner aussi des interférences et des collisions entre les transmissions des dispositifs LoRa.[65] L'efficacité de l'algorithme peut être affectée lorsqu'il est appliqué à des réseaux LoRa à des zones limitées comme dans rayon de 3km.

Et aussi du fait que les nœuds d'un cluster ont la même valeur de SF. Cela entraîne des collisions lorsque les nœuds envoient simultanément des données. Ces résultats reposent sur des hypothèses qui abordent des questions contestables en lien avec l'algorithme de clustering proposé.

L'utilisation de la série Fibonacci pour 500 nœuds						
Facteur d'étalement	SF7	SF8	SF9	SF10	SF11	SF12
Valeurs des k clusters pour les itérations k-means.		5	8	13	21	34
Nombre de nœuds dans chaque bande	29	35	79	107	124	129
La moyenne du nombre nœuds des clusters dans chaque bande		7	10	8	6	4

Tableau 7 : Utilisation de la série Fibonacci pour une valeur de k, le nombre de nœuds et la moyenne

Dans ce tableau nous avons présenté la moyenne du nombre nœuds des clusters dans chaque bande. On constate que dans le cas des SF7, il n'y a pas de cluster car les nœuds les plus proches de la passerelle et qui se situent à l'extérieur de la bande des SF8 sont affectés par des SF7. On constate qu'un nombre de nœud moyenne très petits sur certains clusters au niveau des bandes, interférences.

Les résultats de l'article montrent que l'allocation des facteurs d'étalement aux nœuds des clusters de façon que les clusters plus loin ont des SF grand. Et que ces derniers peuvent atteindre la passerelle sans aucun problème. Comme prévu, la distance entre le périphérique final et la passerelle a une influence considérable sur la probabilité de couverture. Dans le cas de la série de Fibonacci, le modèle a une meilleure probabilité de succès pour des SF inférieurs par rapport à la série carrée, à la série arithmétique et au tableau de Wythoff.

SF	7	8	9	10	11	12
Probabilité de succès	0,79	0,6	0,5	0,4	0,39	0,39

Tableau 8 : Probabilité de succès pour chaque SF

Le tableau présente les facteurs d'étalement des dispositifs finaux dans diverses bandes de fréquences, accompagnés de leurs probabilités de succès. Il est notable que les dispositifs qui ont un SF de 7 et qui sont plus proches de la passerelle affichent une probabilité de succès élevée, tandis que le dispositif des SF plus grands tombe drastiquement. Les facteurs d'étalement

semblent jouer un rôle crucial dans la performance des dispositifs, avec une tendance à des performances plus stables pour le dispositif le plus proche de la passerelle sur l'ensemble des bandes. Ces observations soulignent l'importance de sélectionner le dispositif en fonction de la bande de fréquence cible pour optimiser les chances de succès. Ainsi, une évaluation approfondie des facteurs d'étalement et des probabilités de succès est recommandée pour déterminer le choix optimal du dispositif en fonction du contexte opérationnel et des exigences spécifiques de la bande de fréquences envisagée. Le tableau ci-dessous montre le nombre de nœuds moyenne dans chaque cluster pour les différentes séries utilisées.

Nombre de nœuds moyenne dans chaque cluster pour les différentes séries				
Série	Fibonacci	arithmétique	carrée	Wythoff
Nombre de cluster	34	34	49	37
Nombre de nœuds				
300	2	2	1	2
500	4	4	2	3
700	5	4	3	5

Tableau 9 : le nombre de nœuds moyenne dans chaque cluster pour les différentes séries utilisées

Mais aussi l'affectation de grand SF aux nœuds augmente la consommation d'énergie des nœuds ce qui causé des nœuds qui meurent très vite.

Ces problématiques comme les interférences soulignent les défis spécifiques auxquels sont confrontés notre article de référence. Des recherches supplémentaires sont nécessaires pour développer des adaptations et des améliorations de l'algorithme k-means appliqué aux réseaux LoRa, afin d'optimiser les performances du clustering dans ces environnements. Dans la suite nous allons faire une synthèse du problématique proposé.

B. Synthèse

L'allocation des facteurs d'étalements (SF) dans le contexte de LoRa est cruciale pour optimiser la communication entre les dispositifs IoT et les passerelles. Les facteurs d'étalements déterminent la manière dont les données sont modulées avant d'être transmises, influençant ainsi la portée de la communication. Un SF élevé offre une meilleure résistance au bruit mais à une cadence de transmission plus lente, tandis qu'un SF plus bas permet des transmissions plus rapides mais avec une portée réduite. L'allocation appropriée des SF dépend des conditions du

réseau, de la distance entre les dispositifs et les passerelles, ainsi que des exigences en matière de débit de données. Dans l'article de références pour une de cluster grand, nous constatons des interférences dans la zone de couverture et cela est dû à l'envoi simultané de tous les clusters dans une même bande vers la passerelle. En synthèse, l'utilisation conjointe de l'algorithme de k-means et de la technologie LoRa avec une allocation judicieuse des facteurs d'étalements offre des opportunités significatives pour l'optimisation des réseaux IoT. Notre contribution portera sur l'étapes suivante : sur une étude pour minimiser les interférences. Les conséquences des interférences entre les facteurs d'étalement sont multiples. Elles peuvent entraîner des collisions de données, des erreurs de décodage, et des baisses significatives de débit de transmission. De plus, dans des environnements où plusieurs réseaux coexistent, la compétition pour l'accès au spectre peut accentuer ces interférences, entraînant une dégradation globale des performances du système. Cette contribution semble apporter une amélioration de l'article de référence. Cette approche contribue éviter les interférences et minimiser la perte de débit et à maximiser l'efficacité énergétique et la fiabilité des communications IoT sur de longues distances.

Conclusion

L'algorithme de k-means demeure l'une des méthodes de clustering les plus utilisées dans le contexte des réseaux LoRa (Long Range). Son efficacité réside dans sa simplicité et sa capacité à partitionner un ensemble de données en clusters compacts, ce qui s'avère particulièrement utile pour regrouper des dispositifs LoRa en fonction de leurs caractéristiques communes. L'algorithme K-means offre des avantages significatifs en termes de rapidité de convergence, ce qui en fait un choix judicieux pour la segmentation des dispositifs LoRa, tout en tenant compte des contraintes de ressources inhérentes à ces réseaux à faible consommation d'énergie.

Ayant examiné l'application de l'algorithme de clustering k-means dans le contexte des réseaux LoRa, nous passons désormais à une perspective proactive en explorant une proposition visant à minimiser les interférences dans ce domaine.

Chapitre 4 : Proposition d'une solution basée sur k-means pour minimiser les interférences dans un réseau LoRaWAN

Introduction

L'algorithme de clustering k-means est une méthode populaire de regroupement de données non étiquetées en groupes homogènes. Lorsqu'il est appliqué dans le contexte des réseaux LoRa (Long Range), il peut être utilisé pour regrouper des dispositifs LoRa en clusters basés sur leurs caractéristiques et leurs interactions.

Cette proposition se concentre sur l'amélioration de l'algorithme étudié en haut en introduisant la notion de zone. L'idée est de minimiser les interférences co-SF mais aussi d'optimiser la consommation énergétique en activant sélectivement des zones spécifiques du réseau en fonction des variations dynamiques des conditions environnementales ou des besoins de collecte de données. Cette stratégie vise à maximiser l'utilisation des ressources, contribuant ainsi à la durabilité globale du réseau.

Dans les sections suivantes, nous détaillons l'approche de notre proposition, illustrons son impact potentiel sur l'efficacité énergétique et la performance du réseau. Nous concluons en discutant des avantages substantiels que cette proposition pourrait apporter aux réseaux LoRa dans des scénarios diversifiés de l'Internet des Objets.

I. Contexte et motivation

Considérez la région de Casamance, située au Sénégal, comme notre domaine d'étude. Cette région revêt une importance capitale en tant que hub agricole prospère, jouant un rôle central dans la sécurité alimentaire nationale. Cependant, l'agriculture dans cette zone est confrontée à divers défis majeurs, notamment la nécessité d'optimiser les techniques d'irrigation, de surveiller la qualité des sols et de prendre des décisions éclairées en temps réel. Pour répondre à ces défis, l'intégration des technologies de l'IoT dans le secteur agricole, en particulier l'agriculture intelligente, se présente comme une solution prometteuse. Cette transition requiert une stratégie de communication sophistiquée afin de garantir un déploiement efficace des dispositifs finaux LoRa. C'est dans cette optique que cette étude propose l'utilisation de l'algorithme k-means dans le cadre du réseau LoRaWAN, afin d'optimiser la communication et de résoudre les problèmes d'interférence potentiels.

La nature de solution de notre contribution concerne trois algorithmes : le premier algorithme concerne le découpage en zone, le deuxième nous permet de générer la liste des zones à activer et le troisième permet d'activer les zones. Dans les sections suivantes, nous allons présenter les étapes de notre méthodologie proposée.

II. Modèle de déploiement et hypothèse du réseau

Nous considérons N nœuds LoRa répartis uniformément dans un réseau LoRaWAN de classe A sans retransmission, utilisé un seul canal sur un rayon de 3 km autour d'une seule passerelle. La figure 22 illustre un déploiement avec $N=500$ sur un rayon $R = 3$ km. La passerelle se trouve au centre et les clusters sont répartis dans chaque bande. Dans chaque bande est alloué une valeur de SF. Dans la zone circulaire, nous l'avons découpé par des zones de même longueur d'angle. Ces zones nous permettent de créer des zones d'activités.

Pour une simulation en vue nous pensons paramétrer les nœuds LoRa comme si : une bande passante BW de 125 kHz, un exposant de perte de trajet η de 2,75, une fréquence de bande de 868 MHz et un rayon de réseau R de 3 km.

La configuration précise de ces paramètres dépend des besoins spécifiques de l'application et des conditions environnementales dans lesquelles le nœud LoRa est déployé.

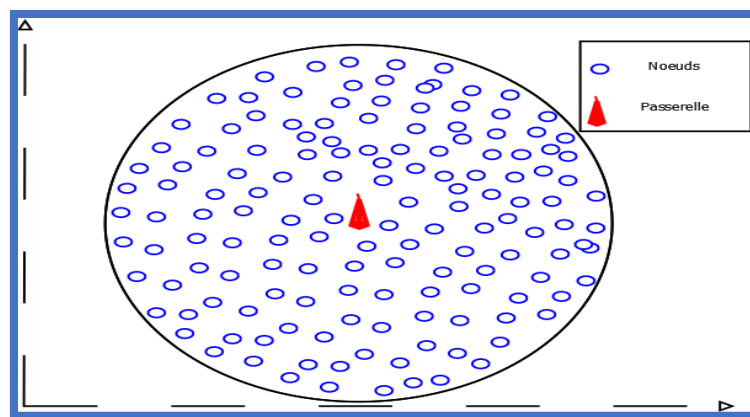


Figure 22: Distribution uniforme de $N=500$ dans une zone de rayon 3 km

III. Etude détaillé de la proposition

La présente proposition vise à résoudre le problème des interférences co-SF (Spread Factor) dans une zone spécifique, en introduisant la notion de zones d'activité. Ces zones sont soigneusement planifiées pour minimiser les interférences et améliorer l'efficacité des communications LoRaWAN. La proposition vise à atténuer ce problème en créant des zones

d'activité bien définies. Dans ce contexte, nous présentons une approche novatrice pour l'activation de zones, s'appuyant sur trois algorithmes synergiques, découpage en zone, liste des zones à activer et enfin l'algorithme de l'activation des listes de zones. Dans cette étude, nous détaillons les étapes clés de nos algorithmes, en mettant en lumière leur complémentarité et leurs avantages respectifs. Nous mettons en évidence son potentiel pour résoudre les problèmes cruciaux liés aux interférences co-SF dans les réseaux LoRaWAN. Cette proposition vise à contribuer significativement à l'efficacité opérationnelle des réseaux LoRaWAN.

A. La Notion de zone

On définit une zone comme étant une superposition verticale de bande de SF différents limité dans l'espace. Elle représente une succession d'arc de cercle avec un angle θ représentant aussi des observations ou des points de données présentant une activité intense, réduite ou nulle. Ce processus est fait pour diminuer les interférences co-SF dans chaque zone. La transmission des données de cluster des nœuds aux têtes de cluster ou bien des têtes de cluster vers la passerelle via LoRa. Elle est définie par un angle délimitant la surface et le nombre de cluster par portion de bande. Une zone peut être active ou non. Une zone est active lorsque les nœuds de la zone sont en mode transmission de données alors qu'elle est inactive lorsque ses nœuds sont en mode sommeil.

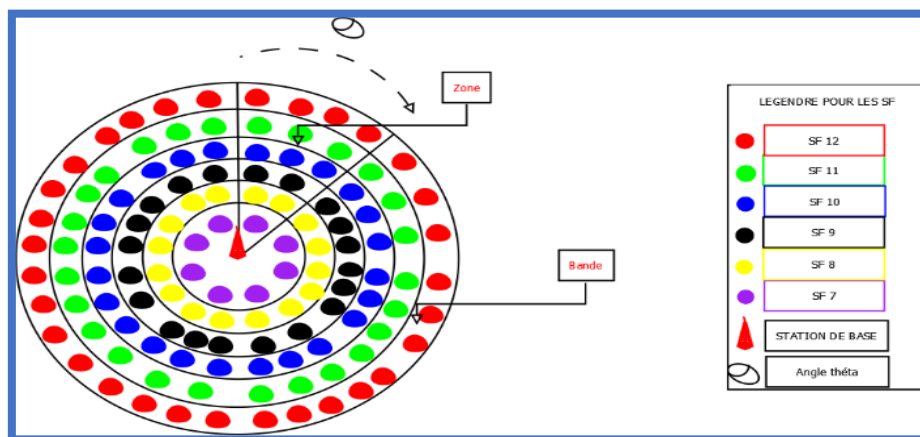


Figure 23: Représentation de zone, de bande et de l'angle θ

Les zones actives ou non dépendent principalement de l'activité des nœuds. Une zone se caractérise par des bandes de cercle contenant des clusters de facteurs d'étalement (FS) différents, couvrant une partie significative de la zone de couverture. L'angle θ associé à ces bandes est généralement optimal, indiquant une étendue considérable de la zone. Dans une zone active, la densité d'événements, telle que la présence de clusters, est élevée, témoignant d'une

région où l'activité est intense ou fréquente c'est-à-dire que les nœuds sont en mode transmission. En revanche, une zone inactive représente un regroupement de clusters où le comportement est considéré comme inexistant c'est-à-dire les nœuds sont en mode sommeil.

B. Découpage en zone

Le principe de découpage en zone fait appel à l'utilisation d'arcs de cercle avec un angle θ pour définir une zone. Cette méthode géométrique permet de découper une zone de couverture définie en plusieurs arcs de cercle. À travers cette approche, nous visons à créer une stratégie de gestion adaptative, permettant une réponse dynamique aux variations d'activité dans des environnements spécifiques. Au cours de cette étude, nous explorerons le principe fondamental de cette méthode de découpage.

Ce processus de découpage en zone est la suivant :

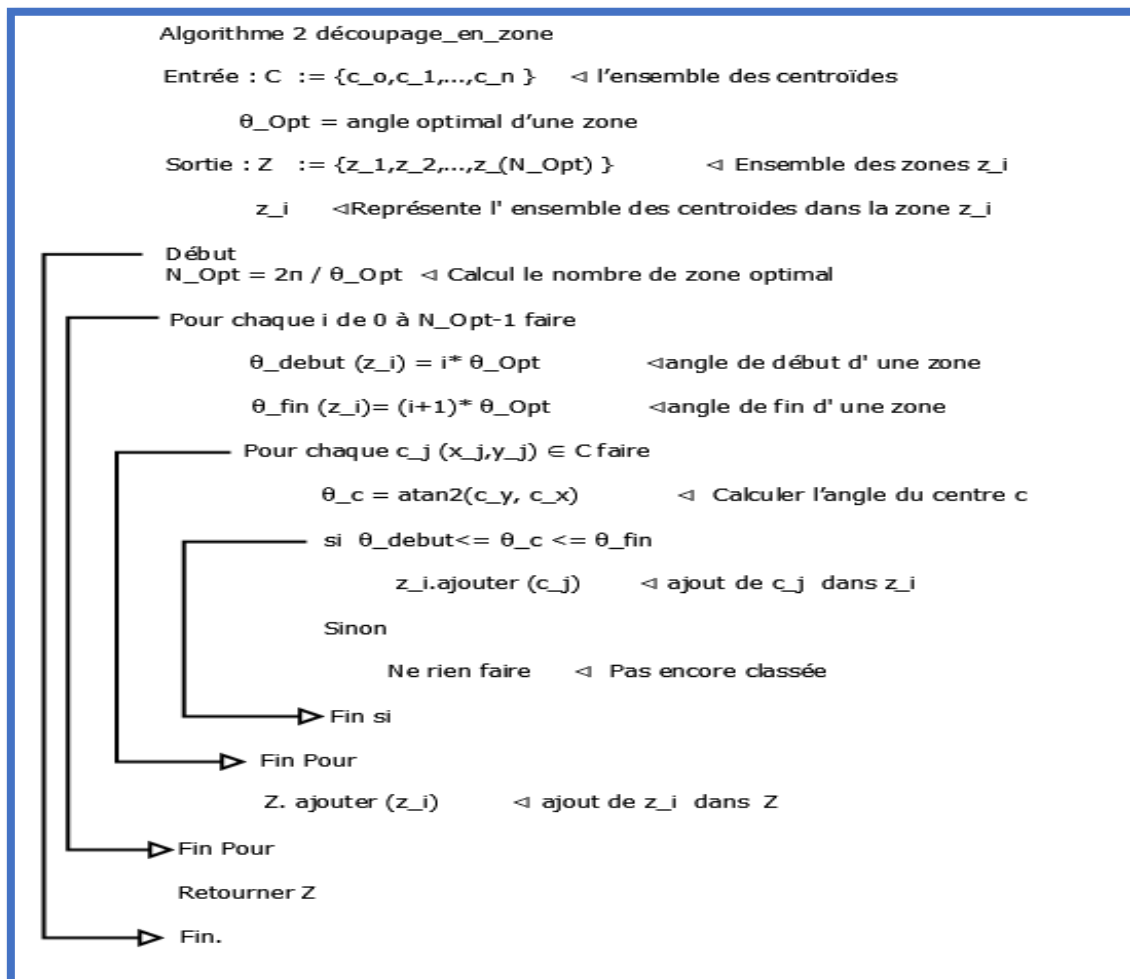


Figure 24: L'algorithme de découpage en zone

L'algorithme "découpage_en_zone" est conçu pour répartir un ensemble de centroïdes dans des zones en fonction d'un angle optimal spécifié. Cet algorithme génère des zones limitées par des angles de départ et de fin pour chaque zone, en utilisant une fonction pour définir la plage d'angles. Les zones sont tracées en fonction de ces bornes d'angle pour chaque cluster. Le but principal de cet algorithme est de regrouper les centroïdes dans des zones en utilisant des divisions angulaires basées sur un angle optimal donné. L'angle optimal (θ_{Opt}) est déterminé par l'utilisateur et représente l'angle qui définit la taille des zones souhaitées. Plus l'angle optimal est petit, plus les zones seront nombreuses et petites. L'algorithme calcule le nombre de zones optimal (N_{Opt}) en divisant 2π degrés par l'angle optimal. Cela garantit que la répartition des zones couvre un cercle complet. Les centroïdes sont classés dans des zones en fonction de leur position angulaire par rapport à l'origine. Chaque centroïde est associé à une zone en fonction de l'angle entre le centroïde et l'axe. Si l'angle du centroïde est inférieur ou égal à l'angle de début de la zone, il est considéré comme déjà classé. Si l'angle du centroïde est compris entre l'angle de début et l'angle de fin de la zone, le centroïde est ajouté à la zone correspondante. Les centroïdes qui ne satisfont aucune des conditions ne sont pas encore classés. L'exécution de cet algorithme génère un grand ensemble appelé l'ensemble des zones Z qui est représenté par le tableau 10 ci-dessous.

L'ENSEMBLE DES ZONES Z			
z_1	z_2	...	$z_{N_{Opt}}$
n_{CH}	n_{CH}	...	n_{CH}

Tableau 10 : Tableau de l'ensemble des zones Z

Ce tableau de l'ensemble Z est utilisé par dans la suite pour l'activation de zone qui est détaillé à la section suivante.

C. Activation de zone

En fonction de l'ensemble Z généré par l'algorithme "découpage_en_zone", les zones d'activité sont planifiées de manière stratégique dans la zone de couverture. Nous utiliserons cet ensemble dans l'algorithme suivant pour la génération des sous-listes de zones. Enfin, nous procédons à l'algorithme d'activation des zones. Cette partie d'activation de zones est scindé en deux algorithmes :

- ❖ Algorithme 3 Listes_des_zones_activées ;
- ❖ Algorithme 4 Activation_zones.

Dans la suite nous vous montrons les processus des deux algorithmes.

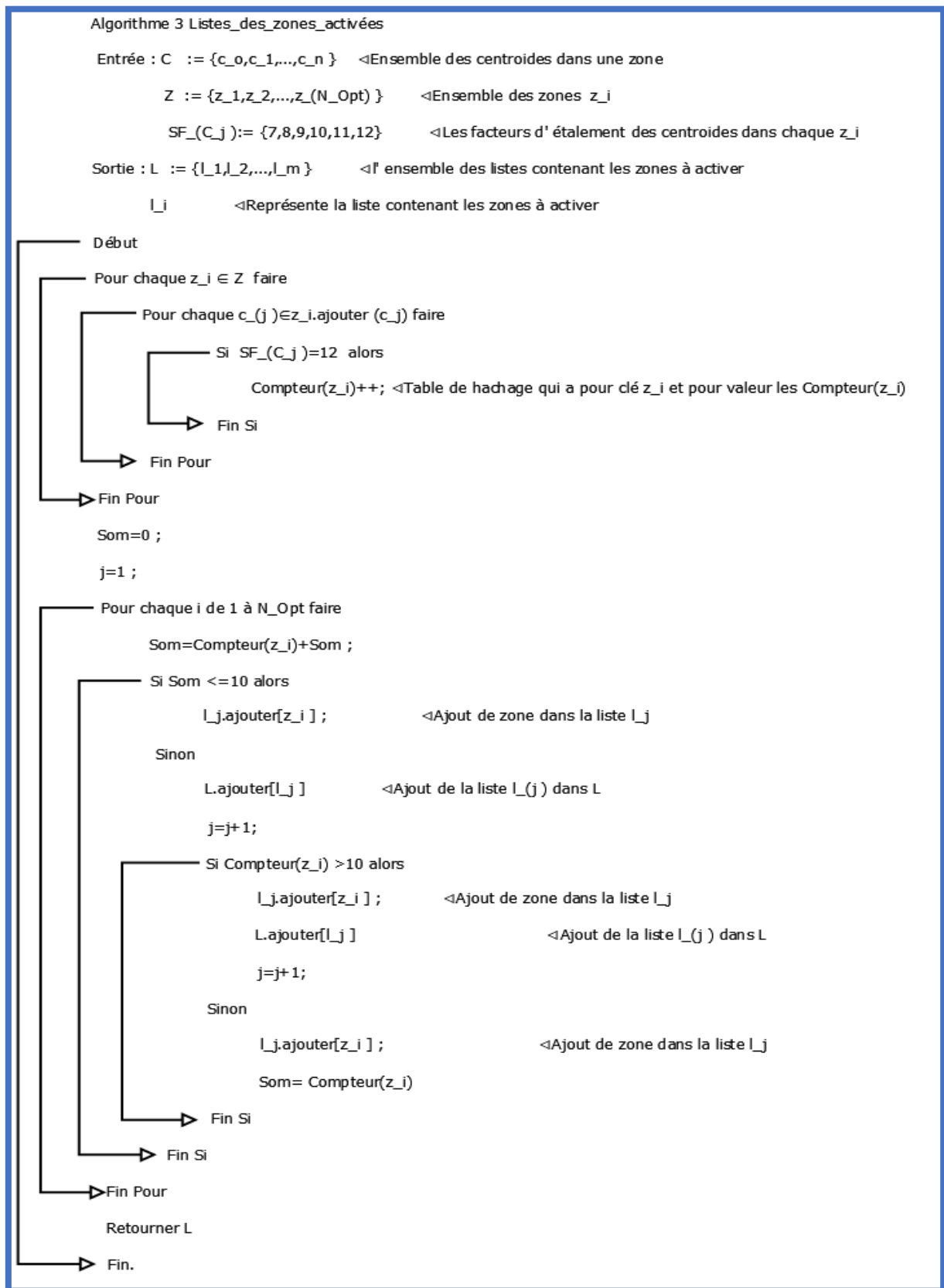


Figure 25: L'algorithme des listes de zones à activer

L'algorithme 3 est conçu pour créer une liste de sous-ensembles de zones à activer, en se basant sur l'évaluation du facteur d'étalement (**SF**) des centroïdes. Il utilise en premier lieu la fonction SF_{C_j} pour évaluer si le facteur d'étalement d'un centroïde est égal à 12. Ces centroïdes sont aussi classés dans un table de hachage qui a pour clé la zone z_i et pour valeur les **Compteur**(z_i).

TABLE DE HACHAGE Compteur (z_i)++			
z_1	z_2	...	z_{N_Opt}
Compteur (z_1)	Compteur (z_2)	...	Compteur (z_{N_Opt})

Tableau 11 : Représentation de la table de hachage

Cette table est utilisée en deuxième lieu pour ajouter les zones dans des sous ensemble l_i de zone à activer. L'ensemble des l_i représente le grand ensemble **L**.

L'ENSEMBLE DES SOUS LISTES L			
l_1	l_2	...	l_j
$l_1.$ <i>ajouter</i> [z_i]	$l_2.$ <i>ajouter</i> [z_i]	...	$l_j.$ <i>ajouter</i> [z_i]

Tableau 12 : Tableau de l'ensemble **L** des sous listes l_i

L'objectif de cet algorithme est de regrouper des zones dans un sous liste l_i . L'ensemble des l_i forme une liste d'ensemble **L**. Cette dernière est dans l'algorithme d'activation de zones pour activer les zones qui sont dans les sous listes l_i . Cet algorithme est représenté en dessous. L'algorithme *activation_zones* permet d'activer les zones de chaque l_i pour une durée déterminée en fonction des paramètres LoRa.

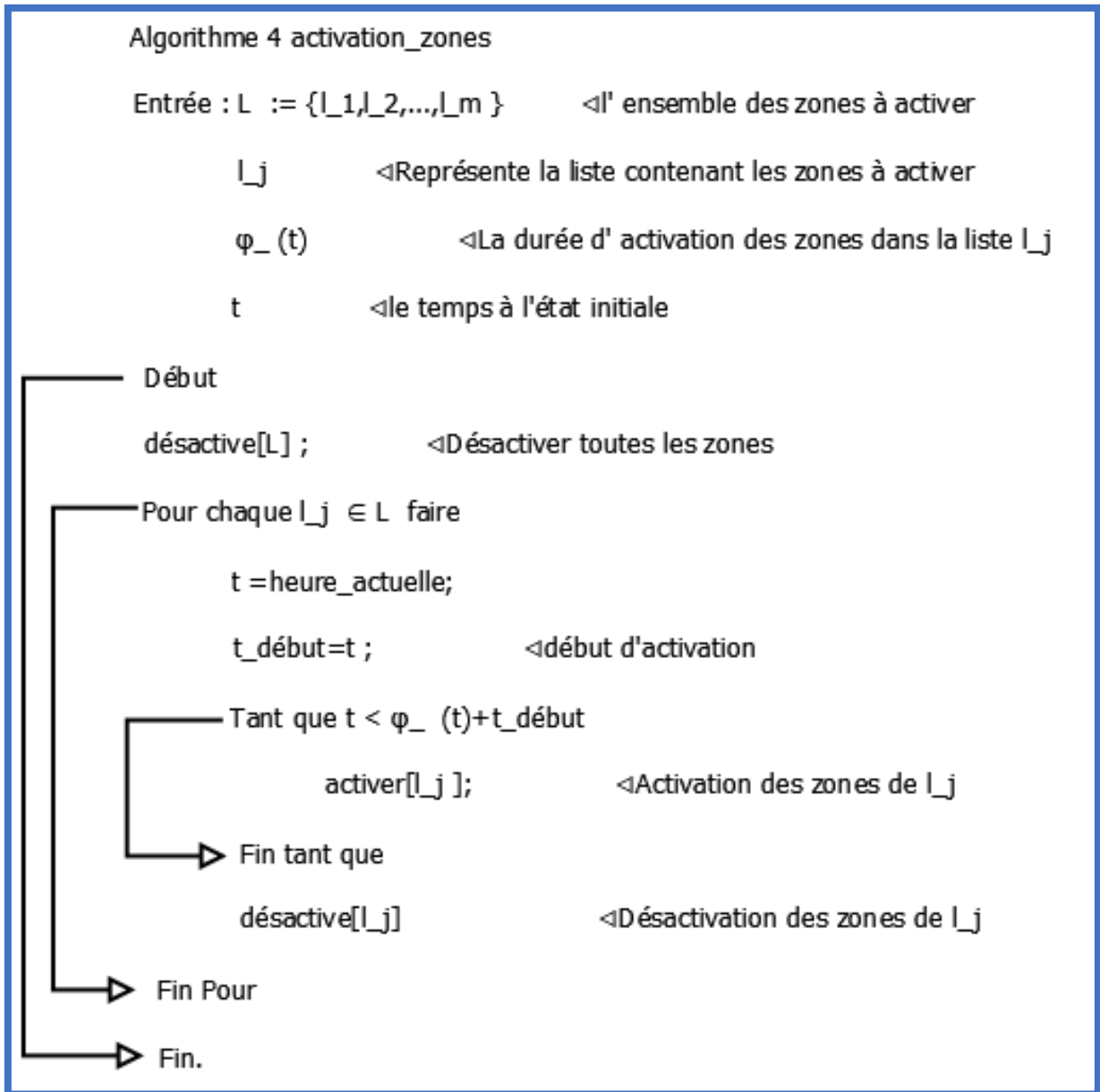


Figure 26: L'algorithme d'activation des listes de zone

L'algorithme 4 est un algorithme de gestion des zones d'activation dans un système, lié à un processus de clustering. L'algorithme gère l'activation et la désactivation de zones basées sur des critères spécifiques liés à des centroïdes dans ces zones. L'algorithme utilise une approche séquentielle pour activer les zones, avec chaque liste l_i représentant un ensemble de zones spécifiques à activer. La fonction $\varphi(t)$ détermine la durée d'activation des zones. L'initialisation et la désactivation initiale assurent un état initial propre avant de commencer le processus d'activation. L'utilisation d'une boucle imbriquée permet d'activer chaque liste l_i pendant la durée spécifiée par $\varphi(t)$, et cela est répété pour chaque liste présente dans l'ensemble L . Après l'activation de toutes les zones, et désactive à nouveau toutes les zones.

L'ENSEMBLE DES SOUS LISTES A ACTIVER			
l_1	l_2	...	l_j
$l_1.ajouter[z_i]$	$l_2.ajouter[z_i]$...	$l_j.ajouter[z_i]$

Tableau 13 : Tableau de l'ensemble des sous listes à activer

En utilisant ces zones d'activités, nous pouvons envoyer les données collectées par les nœuds d'une tête de cluster vers une passerelle située à proximité de la zone. Cela permet de réduire la consommation d'énergie en évitant des transmissions longue distance, diminué des interférences et d'améliorer l'efficacité du réseau LoRa.

IV. Discussion des algorithmes proposés

Les trois algorithmes que nous allons discuter présentent des approches distinctes pour résoudre les problèmes d'interférences dans les zones de couverture. Chacun d'eux répond à des besoins particuliers, que ce soit dans la répartition spatiale de centroïdes ou dans la gestion dynamique des zones d'activation. L'application de cette proposition devrait entraîner une amélioration notable de la qualité des communications dans la zone, conduisant ainsi à une utilisation plus efficace des ressources du réseau et à une augmentation de la fiabilité des transmissions.

Dans la suite nous allons faire une discussion de chacun des trois algorithmes.

A. L'algorithme de découpage en zone

L'algorithme "**découpage_en_zone**" est élaboré pour la répartition optimale de centroïdes dans des zones, basée sur des critères angulaires spécifiques définis par l'utilisateur. En utilisant des angles de début et de fin, cet algorithme divise l'espace autour des centroïdes en zones distinctes. Le paramètre clé est l'angle optimal, influençant la taille et le nombre de zones générées. Applications potentielles incluent la segmentation spatiale et la partition d'objets en fonction de leur position angulaire. L'objectif est de répartir des centroïdes dans des zones en fonction d'un angle optimal défini par l'utilisateur. La méthode est de génère des zones délimitées par des angles de début et de fin, basées sur un angle optimal. Calcule le nombre de zones optimal et classe les centroïdes dans ces zones en fonction de leur position angulaire. Les paramètres sont l'angle optimal (θ_{Opt}) pour définir la taille des zones. Plus N_{Opt} est petit, plus les zones sont nombreuses et petites. Cet algorithme nous permet de découper la zone de couverture en sous zone. Chaque sous zone regroupe un ensemble de CH, ce qui nous permette de savoir le nombre de CH dans chaque zone.

Le découpage de la zone de couverture d'un réseau LoRa en sous-zones présente plusieurs avantages :

Meilleure gestion du réseau : En divisant la zone de couverture en sous-zones, il est plus facile de gérer et de surveiller les dispositifs connectés. Cela permet une meilleure répartition des ressources et une optimisation des performances du réseau.

Réduction des interférences : En limitant la portée de chaque sous-zone, les risques d'interférences entre les dispositifs sont réduits. Cela peut améliorer la qualité et la fiabilité des communications au sein du réseau.

Évolutivité améliorée : En créant des sous-zones, il est plus simple d'ajouter de nouveaux dispositifs au réseau sans compromettre sa performance globale. Cela permet une évolutivité plus fluide et une expansion facilitée du réseau.

Sécurité accrue : Avec un découpage en sous-zones, il est possible de mieux contrôler l'accès aux données et aux dispositifs connectés. Cela renforce la sécurité du réseau en limitant les points d'entrée potentiels pour les cyberattaques.

La complexité temporelle dépend du nombre de centroïdes, du nombre de zones optimal.

$$O(n \times N_{Opt})$$

Un nombre élevé de centroïdes et un angle optimal petit peuvent augmenter les ressources nécessaires.

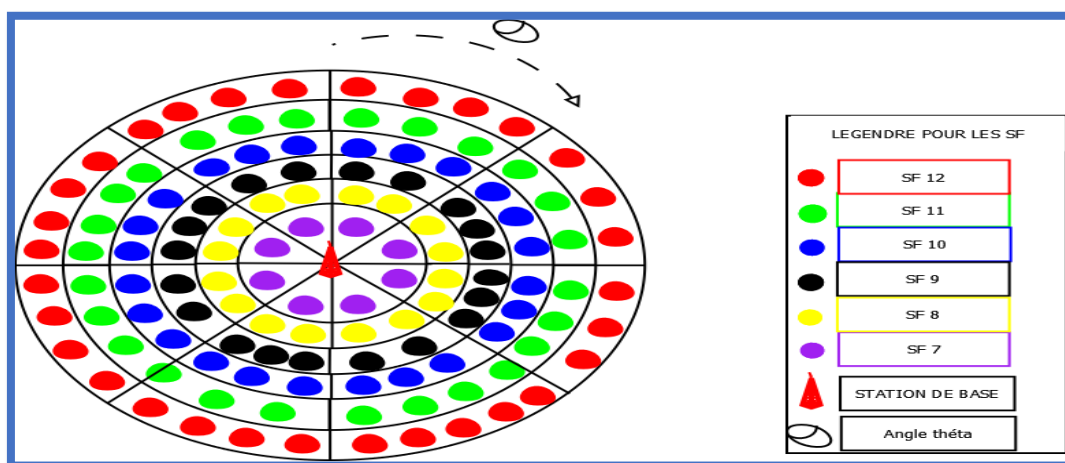


Figure 27: Représentation des zones

B. L'algorithme de liste des zones à activer

L'algorithme " **Listes_des_zones_à_activer** " d'évaluation basée sur le facteur d'étalement (**SF**) semble se concentrer sur la sélection de centroïdes en fonction d'un critère spécifique, en l'occurrence, un facteur d'étalement égal à 12. Les CH qui satisfont cette condition sont classés dans une table de hachage par zone, et des sous-ensembles de zones sont ajoutés à une liste d'activation. L'algorithme pourrait être intégré dans des systèmes nécessitant une évaluation fine des CHs basée sur des critères spécifiques.

L'ajout des sous-ensembles de zones à activer (I_i) à la liste **L** indique une approche séquentielle pour l'activation des zones.

La complexité temporelle de l'algorithme dépendra de facteurs tels que le nombre de CHs, le nombre de zones.

$$O(N_{Opt}^2)$$

L'objectif global des deux algorithmes semble être de gérer la création et l'activation de sous-ensembles de zones en fonction de certaines conditions liées aux centroïdes.

C. L'algorithme d'activation de zone

L'algorithme de gestion des zones d'activation prend une approche dynamique en activant et désactivant des zones spécifiques en fonction de critères liés aux CHs. L'évaluation du facteur d'étalement des nœuds guide la sélection de zones à activer, organisées dans une structure de table de hachage. Le processus d'activation est séquentiel, avec des listes définies pour chaque ensemble de zones à activer. Cet algorithme semble adapté à des scénarios où la gestion dynamique des zones est cruciale. L'objectif est de gérer l'activation et la désactivation des listes de zones basées sur des critères spécifiques liés à des CHs. Cette méthode active séquentiellement les zones en fonction de listes spécifiques I_i avec une durée déterminée par la fonction $\varphi(t)$. Les paramètres sont le facteur d'étalement des nœuds, listes d'activation (I_i), fonction $\varphi(t)$ pour la durée d'activation. Concernant $\varphi(t)$ qui détermine la durée d'activation des zones, dans les réseaux LoRa utilisent différents facteurs de propagation (Spreading Factors - SF) pour moduler les signaux. Des SF plus élevés augmentent la portée mais diminuent le débit de données. Le choix de la bande passante (Bandwidth) et du taux de codage (Coding Rate) influence également la vitesse de transmission des données.

D'Activer certaines zones et laisser les autres en sommeil dans un réseau LoRa peut présenter plusieurs avantages stratégiques et opérationnels :

Économie d'énergie : En laissant certaines zones en sommeil, les dispositifs dans ces zones peuvent économiser de l'énergie en ne transmettant pas de données. Cela prolonge la durée de vie de la batterie des appareils connectés.

Optimisation des ressources : En activant certaines zones prioritaires et en mettant d'autres en sommeil, il est possible de concentrer les ressources du réseau là où elles sont le plus nécessaires. Cela peut améliorer l'efficacité globale du réseau et garantir une utilisation optimale des ressources disponibles.

Gestion flexible du trafic : En activant ou en mettant en sommeil des zones spécifiques en fonction de la demande ou de la charge de travail, il est possible de gérer plus efficacement le trafic et d'éviter les engorgements potentiels. Cela contribue à maintenir des performances constantes et fiables du réseau.

La complexité temporelle dépend du nombre de centroïdes, du nombre de zones optimal, et de la durée d'activation spécifiée.

$$O(N_{Opt})$$

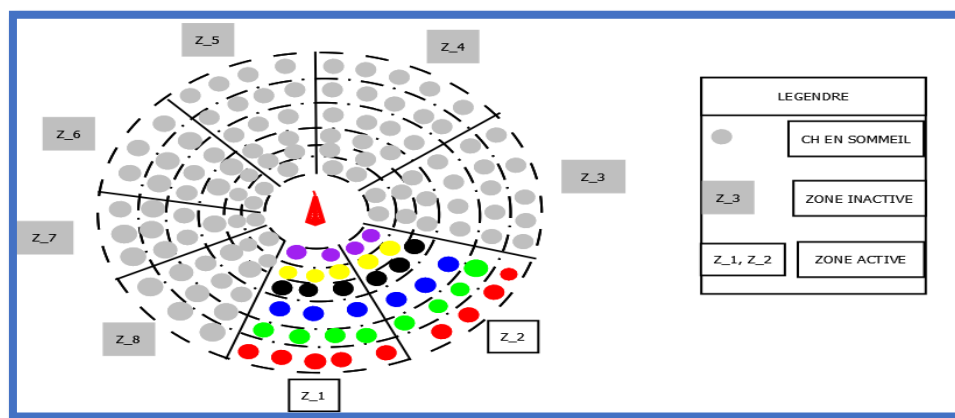


Figure 28: Représentation de deux zones activées

Ensemble, ces algorithmes offrent des solutions variées pour des applications diverses, de la répartition spatiale des nœuds à la gestion dynamique des zones d'activation. Chacun d'eux apporte une contribution unique à la résolution de problèmes spécifiques liés à la minimisation des interférences co-SF dans une zone de couverture. Cette approche peut améliorer la consommation d'énergie globale du réseau LoRa en évitant des transmissions de longue distance et en regroupant les nœuds d'un même cluster. Cela peut également réduire les temps

de latence et augmenter la fiabilité de la transmission des données mais aussi d'augmenter la probabilité de succès dans le réseau. L'avantage de notre proposition est de : si les interférences étaient de 80% dans l'article de références. L'application de cette contribution peut minimiser les interférences à une valeur près de 60%. Cette approximative théorique nous permettrons de dire que notre contribution est une solution qui nous aide à minimiser les interférences co-SF dans les réseaux LoRa.

Il est important de noter que la combinaison de k-means et LoRa présente également certains défis potentiels. Par exemple, la latence des communications LoRa peut limiter la vitesse de transmission des données et le temps de réponse. De plus, la faible capacité de bande passante de LoRa peut poser des problèmes si les ensembles de données sont volumineux. Cette proposition présente d'autres limites :

Complexité de gestion : La gestion d'un réseau LoRa avec plusieurs zones actives et dormantes peut être complexe, nécessitant une infrastructure de contrôle et de gestion sophistiquée pour coordonner efficacement les changements d'état des zones.

Coût potentiel : La mise en place d'une infrastructure permettant de gérer les différentes zones du réseau LoRa peut entraîner des coûts supplémentaires en termes de matériel, de logiciels et de maintenance.

Temps de réaction : Activer ou mettre en sommeil des zones du réseau LoRa peut nécessiter un certain temps de réaction, ce qui peut être un inconvénient si des changements rapides sont nécessaires en fonction des conditions du réseau.

Cependant, en optimisant les paramètres de configuration et en concevant des mécanismes de transmission de données efficaces, ces défis peuvent être atténués.

Conclusion

L'algorithme de découpage en zone offre une base solide en identifiant des zones homogènes, facilitant ainsi la planification stratégique des zones d'activité. Cet ensemble Z généré devient ensuite la pierre angulaire de l'algorithme de génération de sous-listes, où chaque sous-liste est optimisée pour minimiser les interférences spécifiques à sa zone respective. Cette approche élaborée permet une utilisation plus intelligente du spectre radioélectrique, optimisant ainsi la performance globale du réseau L'algorithme d'activation des zones, en utilisant les paramètres SF définis par l'algorithme de génération de sous-listes, permet une mise en œuvre séquentielle

et dynamique des configurations SF. Cette adaptabilité en temps réel assure une réduction significative des interférences co-SF, améliorant ainsi la stabilité des communications LoRaWAN dans la zone cible.

Cette proposition offre une perspective prometteuse pour l'optimisation des performances des réseaux LoRaWAN, contribuant ainsi à l'avancement continu des technologies de communication sans fil dans des contextes sujets aux défis complexes des interférences co-SF.

Des simulations informatiques seront nécessaire pour évaluer les performances de la proposition. Les paramètres tels que le taux de succès de la transmission, le débit de données et la latence.

Conclusion générale et perspectives

L'approche basée sur k-means appliqué dans LoRa et l'utilisation de zones d'activités offre également des avantages en termes de planification du réseau et d'optimisation des ressources. En regroupant les capteurs avec des niveaux de signal similaires, nous pouvons identifier les zones de couverture avec des niveaux de signal élevés et faibles, ce qui peut guider la planification du déploiement des capteurs et l'optimisation de la couverture réseau.

Cette approche pour l'envoi des données des têtes de clusters vers la passerelle offre des avantages significatifs en termes d'optimisation de la consommation d'énergie, de gestion des ressources réseau et d'amélioration de la qualité de la transmission des données. Cela ouvre également la voie à de nouvelles possibilités d'optimisation et de planification du réseau LoRa dans des environnements réels.

La résolution des problèmes des interférences entre les facteurs d'étalement dans les réseaux LoRa ouvre la voie à des déploiements IoT plus robustes, économes en énergie et capables de répondre aux exigences croissantes de connectivité dans un monde de plus en plus interconnecté. Cela renforce la pertinence et l'efficacité des réseaux LoRa dans une multitude d'applications, allant de la gestion intelligente des ressources à la ville intelligente, contribuant ainsi à façonner un avenir plus connecté et durable.

Cependant, il convient de noter que la mise en œuvre de cette solution nécessite une évaluation approfondie et des tests dans des environnements réels. Des facteurs tels que la densité des dispositifs IoT, les contraintes de puissance et les caractéristiques du réseau doivent être pris en compte pour adapter la proposition à des scénarios spécifiques.

Ce travail ouvre la voie à plusieurs autres pistes de recherche à explorer. Dans le cadre de nos travaux futurs, nous nous orientons d'une part vers l'amélioration de l'approche proposée :

- intra et inter clustering ;
- suppression de certain cluster qui ont moins de nœuds ;
- augmenter la zone de couverture.

D'autre part, nous pourrions contribuer à améliorer notre compréhension des algorithmes et à développer des outils plus efficaces pour évaluer leur performance dans une variété de contextes d'application en utilisant des méthodes mathématiques pour évaluer les algorithmes. Cela peut contribuer à améliorer la robustesse du réseau contre les attaques.

Bibliographie

- [1] K. E. Nolan, M. Y. Kelly, M. Nolan, J. Brady, et W. Guibene, « Techniques for resilient real-world IoT », in *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, sept. 2016, p. 222-226. doi: 10.1109/IWCMC.2016.7577061.
- [2] Research Scholar, Electronics & Communication, RGPV Bhopal, MP, India., A. Khandelwal*, Y. K. Jain, et Associate Professor Department of Electronics and Instrumentation Engineering, SATI Vidisha, MP, India., « Cluster Head Selection and Packet Delivery Estimation Based on K-Means and FCM », *Int. J. Recent Technol. Eng. IJRTE*, vol. 8, n° 4, p. 10028-10036, nov. 2019, doi: 10.35940/ijrte.D9231.118419.
- [3] J. Finnegan et S. Brown, « A Comparative Survey of LPWA Networking ». arXiv, 12 février 2018. Consulté le: 13 janvier 2023. [En ligne]. Disponible sur: <http://arxiv.org/abs/1802.04222>
- [4] M. S. Sim, Y.-G. Lim, S. H. Park, L. Dai, et C.-B. Chae, « Deep Learning-Based mmWave Beam Selection for 5G NR/6G With Sub-6 GHz Channel Information: Algorithms and Prototype Validation », *IEEE Access*, vol. 8, p. 51634-51646, 2020, doi: 10.1109/ACCESS.2020.2980285.
- [5] M. S. Batta, H. Mabed, Z. Aliouat, et S. Harous, « A Distributed Multi-Hop Intra-Clustering Approach Based on Neighbors Two-Hop Connectivity for IoT Networks », *Sensors*, vol. 21, n° 3, p. 873, janv. 2021, doi: 10.3390/s21030873.
- [6] R. Molina-Masegosa, J. Gozalvez, et M. Sepulcre, « Comparison of IEEE 802.11p and LTE-V2X: An Evaluation With Periodic and Aperiodic Messages of Constant and Variable Size », *IEEE Access*, vol. 8, p. 121526-121548, 2020, doi: 10.1109/ACCESS.2020.3007115.
- [7] National Institute of Technology, Srinagar, 190006, India, S. Kaur, et R. Naaz Mir, « Clustering in Wireless Sensor Networks- A Survey », *Int. J. Comput. Netw. Inf. Secur.*, vol. 8, n° 6, p. 38-51, juin 2016, doi: 10.5815/ijcnis.2016.06.05.
- [8] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, et M. Ayyash, « Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications », *IEEE Commun. Surv. Tutor.*, vol. 17, n° 4, p. 2347-2376, 2015, doi: 10.1109/COMST.2015.2444095.
- [9] S. Meka et B. Fonseca, « Improving Route Selections in ZigBee Wireless Sensor Networks », *Sensors*, vol. 20, n° 1, p. 164, déc. 2019, doi: 10.3390/s20010164.

- [10] A. I. Maass, D. Nešić, R. Postoyan, et P. M. Dower, « L_p stability of networked control systems implemented on WirelessHART », *Automatica*, vol. 109, p. 108514, nov. 2019, doi: 10.1016/j.automatica.2019.108514.
- [11] M. Bouaziz et A. Rachedi, « A survey on mobility management protocols in Wireless Sensor Networks based on 6LoWPAN technology », *Comput. Commun.*, vol. 74, p. 3-15, janv. 2016, doi: 10.1016/j.comcom.2014.10.004.
- [12] M. B. Yassein, W. Mardini, et A. Khalil, « Smart homes automation using Z-wave protocol », in *2016 International Conference on Engineering & MIS (ICEMIS)*, sept. 2016, p. 1-6. doi: 10.1109/ICEMIS.2016.7745306.
- [13] R. S. Sinha, Y. Wei, et S.-H. Hwang, « A survey on LPWA technology: LoRa and NB-IoT », *ICT Express*, vol. 3, n° 1, p. 14-21, mars 2017, doi: 10.1016/j.icte.2017.03.004.
- [14] S. Darroudi et C. Gomez, « Bluetooth Low Energy Mesh Networks: A Survey », *Sensors*, vol. 17, n° 7, p. 1467, juin 2017, doi: 10.3390/s17071467.
- [15] M. Chiani et A. Elzanaty, « On the LoRa Modulation for IoT: Waveform Properties and Spectral Analysis », *IEEE Internet Things J.*, vol. 6, n° 5, p. 8463-8470, oct. 2019, doi: 10.1109/JIOT.2019.2919151.
- [16] L. F. Livi et J. Catani, « A New Remote Monitor and Control System Based on Sigfox IoT Network ». arXiv, 3 septembre 2021. Consulté le: 12 janvier 2023. [En ligne]. Disponible sur: <http://arxiv.org/abs/2110.15208>
- [17] M. Chen, Y. Miao, Y. Hao, et K. Hwang, « Narrow Band Internet of Things », *IEEE Access*, vol. 5, p. 20557-20577, 2017, doi: 10.1109/ACCESS.2017.2751586.
- [18] M. A. Ertürk, M. A. Aydın, M. T. Büyükakkaşlar, et H. Evirgen, « A Survey on LoRaWAN Architecture, Protocol and Technologies », *Future Internet*, vol. 11, n° 10, p. 216, oct. 2019, doi: 10.3390/fi11100216.
- [19] P. Pandey, « Comparative study of long range communications systems for IoT-Cellular, LoRA & Sigfox », 2018, doi: 10.13140/RG.2.2.32516.01927.
- [20] A. D. Zayas et P. Merino, « The 3GPP NB-IoT system architecture for the Internet of Things », in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, mai 2017, p. 277-282. doi: 10.1109/ICCW.2017.7962670.

- [21] G. Strazdins, A. Elsts, K. Nesenbergs, et L. Selavo, « Wireless Sensor Network Operating System Design Rules Based on Real-World Deployment Survey », *J. Sens. Actuator Netw.*, vol. 2, n° 3, p. 509-556, août 2013, doi: 10.3390/jsan2030509.
- [22] R. Piyare, A. Murphy, M. Magno, et L. Benini, « On-Demand LoRa: Asynchronous TDMA for Energy Efficient and Low Latency Communication in IoT », *Sensors*, vol. 18, n° 11, p. 3718, nov. 2018, doi: 10.3390/s18113718.
- [23] A. Jebril, A. Sali, A. Ismail, et M. Rasid, « Overcoming Limitations of LoRa Physical Layer in Image Transmission », *Sensors*, vol. 18, n° 10, p. 3257, sept. 2018, doi: 10.3390/s18103257.
- [24] H. Wang et A. O. Fapojuwo, « A Survey of Enabling Technologies of Low Power and Long Range Machine-to-Machine Communications », *IEEE Commun. Surv. Tutor.*, vol. 19, n° 4, p. 2621-2639, 2017, doi: 10.1109/COMST.2017.2721379.
- [25] Z. Qin, Y. Liu, et J. A. McCann, « Performance Analysis of Clustered LoRa Networks », p. 13.
- [26] L. Vangelista, « Frequency Shift Chirp Modulation: The LoRa Modulation », *IEEE Signal Process. Lett.*, vol. 24, n° 12, p. 1818-1821, déc. 2017, doi: 10.1109/LSP.2017.2762960.
- [27] D. Garlisi, A. Martino, J. Zouwayhed, R. Pourrahim, et F. Cuomo, « Exploratory approach for network behavior clustering in LoRaWAN », *J. Ambient Intell. Humaniz. Comput.*, mars 2021, doi: 10.1007/s12652-021-03121-z.
- [28] R. Yasmin, J. Petajajarvi, K. Mikhaylov, et A. Pouttu, « On the integration of LoRaWAN with the 5G test network », in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Montreal, QC: IEEE, oct. 2017, p. 1-6. doi: 10.1109/PIMRC.2017.8292557.
- [29] N. Blenn et F. Kuipers, « LoRaWAN in the Wild: Measurements from The Things Network ». arXiv, 9 juin 2017. Consulté le: 11 août 2022. [En ligne]. Disponible sur: <http://arxiv.org/abs/1706.03086>
- [30] M. S. A. Muthanna, A. Muthanna, A. Rafiq, M. Hammoudeh, S. Lynch, et A. A. A. El-Latif, « Deep Reinforcement Learning based Transmission Policy Enforcement and Multi-Hop Routing in QoS aware LoRa IoT Networks ».

- [31] M. S. A. Muthanna, P. Wang, M. Wei, A. Rafiq, et N. N. Josbert, « Clustering Optimization of LoRa Networks for Perturbed Ultra-Dense IoT Networks », *Information*, vol. 12, n° 2, p. 76, févr. 2021, doi: 10.3390/info12020076.
- [32] « LoRa Alliance® ». Consulté le: 11 août 2022. [En ligne]. Disponible sur: <https://loralliance.org/resource-hub/lorawanr-specification-v11>
- [33] A. A. Abbasi et M. Younis, « A survey on clustering algorithms for wireless sensor networks », *Comput. Commun.*, vol. 30, n° 14, p. 2826-2841, oct. 2007, doi: 10.1016/j.comcom.2007.05.024.
- [34] M. B. Mebarek, « La proposition d'une approche de l'extraction d'informations basée sur l'algorithme K-Means ».
- [35] A. Stetco *et al.*, « Machine learning methods for wind turbine condition monitoring: A review », *Renew. Energy*, vol. 133, p. 620-635, avr. 2019, doi: 10.1016/j.renene.2018.10.047.
- [36] J. Confais et M. L. Guen, « Premiers pas en régression linéaire avec SAS® ».
- [37] B. Mahesh, « Machine Learning Algorithms - A Review », vol. 9, n° 1, 2018.
- [38] M. B. Mebarek, « La proposition d'une approche de l'extraction d'informations basée sur l'algorithme K-Means ».
- [39] S. Arjunan et S. Pothula, « A survey on unequal clustering protocols in Wireless Sensor Networks », *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 31, n° 3, p. 304-317, juill. 2019, doi: 10.1016/j.jksuci.2017.03.006.
- [40] M. Aldeer, R. Howard, et A. Al-Hilli, « Minimizing energy consumption in transmit-only sensor networks via optimal placement of the cluster heads », in *Proceedings of the Eighth Wireless of the Students, by the Students, and for the Students Workshop*, New York City New York: ACM, oct. 2016, p. 36-38. doi: 10.1145/2987354.2987367.
- [41] J. Corn et J. W. Bruce, « Clustering algorithm for improved network lifetime of mobile wireless sensor networks », in *2017 International Conference on Computing, Networking and Communications (ICNC)*, janv. 2017, p. 1063-1067. doi: 10.1109/ICCNC.2017.7876282.
- [42] « La maximisation de la durée de vie des réseaux de capteurs sans fil à base de Fuzzy CMeans.pdf ».
- [43] « Les techniques de clustering dédié aux données multimédia.pdf ».

- [44] W. R. Heinzelman, A. Chandrakasan, et H. Balakrishnan, « Energy-efficient communication protocol for wireless microsensor networks », in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, Maui, HI, USA: IEEE Comput. Soc, 2000, p. 10. doi: 10.1109/HICSS.2000.926982.
- [45] A. Ben Ammar, A. Dziri, M. Terre, et H. Youssef, « Multi-hop LEACH based cross-layer design for large scale wireless sensor networks », in *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, sept. 2016, p. 763-768. doi: 10.1109/IWCMC.2016.7577153.
- [46] L. Hawarah, « Une approche probabiliste pour le classement d'objets incomplètement connus dans un arbre de décision ».
- [47] L. Xu, R. Collier, et G. M. P. O'Hare, « A Survey of Clustering Techniques in WSNs and Consideration of the Challenges of Applying Such to 5G IoT Scenarios », *IEEE Internet Things J.*, vol. 4, n° 5, p. 1229-1249, oct. 2017, doi: 10.1109/JIOT.2017.2726014.
- [48] « Adaptation de clustering phase de protocole de routage LEACH pour les réseaux de capteurs sans fil .pdf ».
- [49] J. Albath, M. Thakur, et S. Madria, « Energy constraint clustering algorithms for wireless sensor networks », *Ad Hoc Netw.*, vol. 11, n° 8, p. 2512-2525, nov. 2013, doi: 10.1016/j.adhoc.2013.05.016.
- [50] K. P. Sinaga et M.-S. Yang, « Unsupervised K-Means Clustering Algorithm », *IEEE Access*, vol. 8, p. 80716-80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [51] Geon Yong Park, Heeseong Kim, Hwi Woon Jeong, et Hee Yong Youn, « A Novel Cluster Head Selection Method based on K-Means Algorithm for Energy Efficient Wireless Sensor Network », in *2013 27th International Conference on Advanced Information Networking and Applications Workshops*, Barcelona: IEEE, mars 2013, p. 910-915. doi: 10.1109/WAINA.2013.123.
- [52] S. Yousefi, F. Derakhshan, H. S. Aghdasi, et H. Karimipour, « An energy-efficient artificial bee colony-based clustering in the internet of things », *Comput. Electr. Eng.*, vol. 86, p. 106733, sept. 2020, doi: 10.1016/j.compeleceng.2020.106733.
- [53] J. C. Bezdek, R. Ehrlich, et W. Full, « FCM: The fuzzy c-means clustering algorithm », *Comput. Geosci.*, vol. 10, n° 2-3, p. 191-203, janv. 1984, doi: 10.1016/0098-3004(84)90020-7.

- [54] A. Djoudi, R. Zitouni, N. Zangar, et L. George, « Reconfiguration of LoRa Networks Parameters using Fuzzy C-Means Clustering », in *2020 International Symposium on Networks, Computers and Communications (ISNCC)*, Montreal, QC, Canada: IEEE, oct. 2020, p. 1-6. doi: 10.1109/ISNCC49221.2020.9297284.
- [55] H. Laguel, « Déploiement sur une plate-forme de visualisation d'un algorithme coopératif pour la segmentation d'images IRM autour d'un système multi-agents ».
- [56] « ETUDE ET IMPLEMENTATION DE QUELQUES VARIANTES DU MODELE FCM POUR LA SEGMENTATION FLOUE DES IMAGES.pdf ».
- [57] P. Kaur, D. A. K. Soni, et D. A. Gosain, « Novel Intuitionistic Fuzzy C-Means Clustering for Linearly and Nonlinearly Separable Data », vol. 11, n° 3, 2012.
- [58] « Analysing Multi-hop Intra-Cluster Communication in Cluster-Based Wireless Sensor Networks.pdf ».
- [59] M. Z. Mammeri, U. P. Sabatier, et M. J. Rodrigues, « Routage pour la Gestion de l'Energie dans les Réseaux de Capteurs Sans Fil ».
- [60] Y. Sung, S. Lee, et M. Lee, « A Multi-Hop Clustering Mechanism for Scalable IoT Networks », *Sensors*, vol. 18, n° 4, p. 961, mars 2018, doi: 10.3390/s18040961.
- [61] V. Sucasas, A. Radwan, H. Marques, J. Rodriguez, S. Vahid, et R. Tafazolli, « A survey on clustering techniques for cooperative wireless networks », *Ad Hoc Netw.*, vol. 47, p. 53-81, sept. 2016, doi: 10.1016/j.adhoc.2016.04.008.
- [62] S. K. Singh, P. Kumar, J. P. Singh, et M. A. A. Alryalat, « An energy efficient routing using multi-hop intra clustering technique in WSNs », in *TENCON 2017 - 2017 IEEE Region 10 Conference*, nov. 2017, p. 381-386. doi: 10.1109/TENCON.2017.8227894.
- [63] T. A. Al-Janabi et H. S. Al-Raweshidy, « Optimised clustering algorithm-based centralised architecture for load balancing in IoT network », in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, août 2017, p. 269-274. doi: 10.1109/ISWCS.2017.8108123.
- [64] A. Muqattash et M. Krunz, « A single-channel solution for transmission power control in wireless ad hoc networks », in *Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing*, Roppongi Hills, Tokyo Japan: ACM, mai 2004, p. 210-221. doi: 10.1145/989459.989486.

- [65] B. Reynders, W. Meert, et S. Pollin, « Range and coexistence analysis of long range unlicensed communication », in *2016 23rd International Conference on Telecommunications (ICT)*, Thessaloniki, Greece: IEEE, mai 2016, p. 1-6. doi: 10.1109/ICT.2016.7500415.
- [66] J. M. Valtorta, A. Martino, F. Cuomo, et D. Garlisi, « A Clustering Approach for Profiling LoRaWAN IoT Devices », in *Ambient Intelligence*, I. Chatzigiannakis, B. De Ruyter, et I. Mavrommati, Éd., in *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2019, p. 58-74. doi: 10.1007/978-3-030-34255-5_5.
- [67] M. Asad Ullah, J. Iqbal, A. Hoeller, R. Souza, et H. Alves, « K-Means Spreading Factor Allocation for Large-Scale LoRa Networks », *Sensors*, vol. 19, n° 21, p. 4723, oct. 2019, doi: 10.3390/s19214723.
- [68] F. Cuomo *et al.*, « Towards traffic-oriented spreading factor allocations in LoRaWAN systems », in *2018 17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, Capri: IEEE, juin 2018, p. 1-8. doi: 10.23919/MedHocNet.2018.8407091.
- [69] U. Noreen, A. Bounceur, et L. Clavier, « A study of LoRa low power and wide area network technology », in *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Fez, Morocco: IEEE, mai 2017, p. 1-6. doi: 10.1109/ATSIP.2017.8075570.
- [70] « De Castro Tome et al. - 2019 - Long-Range Low-Power Wireless Networks and Samplin.pdf ».
- [71] A. Hoeller, R. D. Souza, O. L. Alcaraz Lopez, H. Alves, M. De Noronha Neto, et G. Brante, « Analysis and Performance Optimization of LoRa Networks With Time and Antenna Diversity », *IEEE Access*, vol. 6, p. 32820-32829, 2018, doi: 10.1109/ACCESS.2018.2839064.
- [72] G. Zhu, C.-H. Liao, T. Sakdejayont, I.-W. Lai, Y. Narusue, et H. Morikawa, « Improving the Capacity of a Mesh LoRa Network by Spreading-Factor-Based Network Clustering », *IEEE Access*, vol. 7, p. 21584-21596, 2019, doi: 10.1109/ACCESS.2019.2898239.