REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE IBN KHALDOUN - TIARET

MEMOIRE

Présenté à :

FACULTÉ MATHEMATIQUES ET INFORMATIQUE
DÉPARTEMENT D'INFORMATIQUE

Pour l'obtention du diplôme de :

MASTER

Spécialité : Réseaux et Télécommunications

Par :

ZERADNA Rim

CHORFI Imen

Sur le thème

L'utilisation de l'apprentissage automatique pour la détection des attaques Déni de Service (DOS) dans les réseaux de capteurs sans fil

Soutenu publiquement à Tiaret devant le jury composé de :

Mr. Bengheni Abdelmalek	M.C.B	Université IBN-KHALDOUN Tiaret	Examinateur
Mr. BAKAR KHALED	M.A.A	Université IBN-KHALDOUN Tiaret	Encadreur
Mr. BERBER ELMEHDI	M.A.A	Université IBN-KHALDOUN Tiaret	Président

2021 - 2022

Remerciements

REMERCIEMENTS

Nos remerciements vont tout premièrement à Dieu tout puissant pour la volonté, la santé et la patience, qu'il nous a donné durant toutes ces longues années.

Nous tenons également à exprimer nos vifs remerciements à notre encadreur Monsieur **BAkkAR.k** pour nous avoir encadrés tout le long de la réalisation de ce mémoire et qui n'a pas cessé de nous donner ses conseils, orientations et remarques.

Nos remerciements s'adressent aussi aux membres de jury, pour l'honneur qu'ils nous font en acceptant de juger ce travail.

Nos remerciements vont également à tous les professeurs et enseignants du département Informatique qui nous ont suivi au long de notre cycle d'études.

Nous exprimons nos profondes gratitudes à nos parents pour leurs encouragements et leur soutien.

Dédicaces

Dédicaces

Je dédie se travail à :

A la mémoire de **ma mère** et je la remercie pour son amour, ses encouragements et ses sacrifies.

A mon homme T. BILLE pour son grand patience,
Son interminable soutien et encouragement. Sans oublies mabelle-famille

Les deux personnes qui ont été là pour moi le plus **WALID** et **ISSAM**
Tu étais toujours à mes cotes pour soutenir et m'encourager.
Que ce travail traduit ma gratitude et mon affection

A mon père. Pour son affection et la confiance qu'Ill m'a accordé

A mes très chers frères **SADAM HOUCIN**, **TAREK** et mes Belle sœurs **YOSRA, NARIMEN**

A toute ma famille pour leur soutien tout au long de mon parcours universitaire,
Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fuit de votre soutien infaillible,
Merci d'être toujours là pour moi.

ZERADNA RIM

Dédicaces

A mon très cher père ***Mohamed Chorfi***

Tu as toujours été pour moi un exemple du père respectueux, honnête, de la personne méticuleuse, je tiens à honorer l'homme que tu es.
Aucune dédicace ne saurait exprimer l'amour l'estime et le respect que j'ai toujours eu pour toi.
Ce modeste travail est le fruit de tous les sacrifices que tu as déployés pour mon éducation et ma formation. J'implore le tout-puissantpour qu'il t'accorde une bonne santé et une vie longue et heureuse.

A mon très cher maman ***Kentour Khaldia***

Aucune dédicace très chère maman, ne pourrait exprimer la profondeur des sentiments que j'éprouve pour vous, vos sacrifices innombrables et votre dévouement firent pour moi un encouragement. Ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études.
Vous m'avez aidé et soutenu pendant de nombreuses années avec à chaque fois une attention renouvelée.

Je tiens à remercier aussi ***Younes, Amin, Roumaisaa .Mes tantes et mes oncles***

Aucun langage ne saurait exprimer mon respect et ma considération pour votre soutien et encouragements. Je vous dédie ce travail en reconnaissance de l'amour que vous m'offrez quotidiennement et votre bonté exceptionnelle.

Chorfi Imen

Sommaire

Introduction générale 1
I Les réseaux de capteurs sans fil 4
I.1 Introduction 4
I.2 Définition 4
I.2.1 Un Capteur 4
I.2.1.1 Capteur actif 4
I.2.1.2 Capteur passif 5
I.2.2 Réseau de capteurs sans fil 5
I.3 Architecture d'un réseau de capteurs sans fil 5
I.3.1 Les réseaux de capteurs sans fil plats 6
I.3.2 Les réseaux de capteurs sans fil hiérarchiques 7
I.4 Architecture d'un nœud de capteur 7
I.5 Les domaines d'applications des réseaux de capteurs sans fil 9
I.6 La pile protocolaire 11
I.7 Contraintes des réseaux de capteurs sans fil 13
I.8 Conclusion 15
II La sécurité dans les réseaux de capteurs sans fil 17
II.1 Introduction 17
II.2 La sécurité dans le RCSFs 17
II.3 Objectifs de la sécurité dans le RCSFs 17
II.3.1 L'authentification 17
II.3.2 Autorisation 17
II.3.3 Disponibilité 17
II.3.4 Intégrité 18
II.3.5 Confidentialité 18
II.3.6 Contrôle d'accès 18
II.3.7 Le non répudiation 18
II.4 Les attaques contre les réseaux de capteurs sans fil 18
II.4.1 Classifications des attaquants 18
II.4.2 Selon sa position par rapport au réseau 18
II.4.3 Selon sa capacité 19
II.4.4 Selon son nature 19
II.4.4.1 Attaquant passif 19
II.4.4.2 Attaquant actif 20
II.5 Les mécanismes de sécurité dans les RCSFs 25
II.5.1 La cryptographie 25

II.5.1.1 Le chiffrement 25
II.5.1.2 La signature digitale 26
II.5.1.3 La fonction de hachage 27
II.5.2 Système de détection d'intrusion 27
II.6 Conclusion 28
III Utilisation de l'apprentissage automatique pour détecter les intrusions dans les réseaux de capteurs sans fil 30
III.1 Introduction 30
III.2 L'apprentissage automatique 30
III.2.1 Apprentissage supervisé 30
III.2.1.1 Classification 31
III.2.1.2 Régression 31
III.2.2 Apprentissage non supervisé 31
III.3 Les algorithmes d'apprentissage supervisé 31
III.3.1 K-Plus Proches Voisins (K-PPV) 31
III.3.2 Bayes Naïve 32
III.3.3 L'arbre de décision 32
III.3.4 Machine vectorielle de soutien 32
III.3.5 Régression Logistique 32
III.4 Les algorithmes d'apprentissage non supervisé 32
III.4.1 K-Means 32
III.4.2 Association Rules 33
III.5 WEKA 33
III.5.1 Définition 33
III.5.2 Les métriques d'évaluation 33
III.6 Description « Data Set » 34
III.7 Protocol LEACH… 35
III.8 Expérience et évaluation 36
III.8.1 L'arbre de décision 36
III.8.2 Machine vectorielle de soutien 40
III.8.3 Régression logistique 43
III.8.4 K-Plus Proches Voisins 46
III.8.5 Bayes Naïve 49
III.9 Résultat 51
Conclusion générale 54
Référence bibliothèque 56

Liste des figures

List des figures

Chapitre I

Figure I-1: Schéma simple d'un capteur 4
Figure I-2: schéma simple de réseau de capteurs sans fil 5
Figure I-3: Architecture de réseau de capteurs sans fil 6
Figure I-4: Architecture plats 7
Figure I-5: Architecture hiérarchique 7
Figure I-6: Architecture matérielle d'un capteur 8
Figure I-7: Application militaire 9
Figure I-8: Application médicale 9
Figure I-9: Application environnementales 10
Figure I-10: Application domotique 10
Figure I-11: Application commerciale 11
Figure I-12: Application sécurité 11
Figure I-13: Pile protocolaire dans les RCSF 12

Chapitre II

Figure II-1: Attaque jamming 20
Figure II-2: Attaque de trou noir (Blackhole) 22
Figure II-3 : Attaquant Sybil 22
Figure II-4:Attaque Wormhole 23
Figure II-5: l'attaque Hello 24
Figure II-6: Le chiffrement symétrique 26
Figure II-7 : chiffrement asymétrique 26

Chapitre III

Figure III-1: Weka_ (software) _logo 33
Figure III-2: Le workflow du modèle de classification 36
Figure III-3 : DT Comparaison en terme de nombre de feuille et la taille de l'arbre 39
Figure III-4: DT Comparaison en terme de temps 39
Figure III-5 : DT Comparaison en terme des instances correctement et incorrectement classées 40
Figure III-6: SVM Comparaison en terme de temps 42
Figure III-7: SVM Comparaison en terme des instances correctement et incorrectement classées 43
Figure III-8: LR Comparaison en terme de temps 45
Figure III-9: LR Comparaison en terme des instances correctement et incorrectement classées 45
Figure III-10: K-NN Comparaison en terme de temps 48
Figure III-11: K-NN Comparaison en terme d'instances correctement et incorrectement Classées 48
Figure III-12: naïve bayes Comparaison en terme de temps 51
Figure III-13: comparaison entre les algorithmes en terme des instances correctement classées 52

Liste des tableaux

Chapitre I

……

Chapitre II

……

Chapitre III

Table III-1: Description de l'attribut....35
Table III-2: Résumer des résultats de l'arbre décision obtenus....36
Table III-3: l'exactitude détaillée de l'arbre décision par classe....37
Table III-4: la matrice de confusion d'un l'arbre décision....37
Table III-5: Résumer des résultats de « DT » utilisant des attaques sélectionnée....38
Table III-6: l'exactitude détaillée de « DT » par classe utilisant des attaques sélectionnée....38
Table III-7: la matrice de confusion d'un « DT » utilisant des attaques sélectionnée....38
Table III-8 : Résumer des résultats de la SVM obtenus....41
Table III-9: l'exactitude détaillée de la SVM par classe....41
Table III-10 : la matrice de confusion d'un SVM utilisant des attaques sélectionnée....41
Table III-11: Résumer des résultats de « SVM » utilisant des attaques sélectionnée....41
Table III-12: l'exactitude détaillée de « SVM » par classe utilisant des attaques sélectionnée....42
Table III-13: la matrice de confusion d'un « SVM » utilisant des attaques sélectionnée....42
Table III-14: Résumer des résultats de la LR obtenus....43
Table III-15: l'exactitude détaillée de la LR par classe....44
Table III-16•: la matrice de confusion d'un LR....45
Table III-17: Résumer des résultats de « LR » utilisant des attaques sélectionnée....44
Table III-18: l'exactitude détaillée de « LR » par classe utilisant des attaques sélectionnée....44
Table III-19: la matrice de confusion d'un « LR » utilisant des attaques sélectionnée....45
Table III-20 : Résumer des résultats de la K-NN obtenus....46
Table III-21: l'exactitude détaillée de la K-NN par classe....46
Table III-22: la matrice de confusion d'un K-NN....47
Table III-23: Résumer des résultats de « K-NN » utilisant des attaques sélectionnée....47
Table III-24: l'exactitude détaillée de « K-NN » par classe utilisant des attaques sélectionnée....47
Table III-25: la matrice de confusion d'un « K-NN » utilisant des attaques sélectionnée....48
Table III-26: Résumer des résultats de la Naïve Bayes obtenus....49
Table III-27: l'exactitude détaillée de la Naïve Bayes par classe....49
Table III-28: la matrice de confusion d'un Naïve Bayes....50
Table III-29: Résumer des résultats de «naïve bayes» utilisant des attaques sélectionnée....50
Table III-30: l'exactitude détaillée de « naïve bayes » par classe utilisent des attaques sélectionnée....50
Table III-31: la matrice de confusion d'un «naïve bayes » utilisent des attaques sélectionnée....51

Liste des acronymes

Liste des acronymes :

- **ACL :** Access Control List
- **ADC:** Analog to Digital Converters.
- **ANN**: Artificial Neural Network.
- **BS**: Base Station.
- **CH**: Cluster Head
- **DT**: Decision Trees.
- **DOS** : Denial Of Service
- **EAR :** Eavesdrop And Register.
- **FN**: False Negative
- **FP**: False Positive.
- **FTP :** File Transfer Protocol.
- **GNU** : General Public License
- **HIDS**: Host based Intrusion Detection System
- **HTTP :** Hypertext Transfer Protocol
- **ICMP :** Internet Control Message Protocol
- **IDS :** Intrusion Détection System
- **IP :** Internet Protocol
- **K-NN**: K- Nearest Neighbor.
- **LEACH:** Low-Energy Adaptive Clustering Hierarchy.
- **LR** : Logistique Regression.
- **ML:** Machine Learning
- **NIDS**: Network based Intrusion Detection System
- **RAM:** Random Access Memory.
- **RCSFs :** Réseaux de Capteurs Sans Fil.
- **SAR:** Sequential Assignment Routing.
- **SMACS:** Self-organizing Medium Access Control for Sensors networks

- **SMP:** Sensors Management Protocol.
- **SVM** : Support Vector Machine.
- **TADAP:** Task Assignment and Data Advertisement Protocol.
- **TCP :** Transmission Control Protocol
- **TN**: True Negative.
- **TP**: True Positive
- **UDP :** User Datagram Protocol
- **UDP-Like:** User Datagram Protocol Like.
- **WEKA**: Waikato Environment for Knowledge Analysis.
- **WSN**: Wireless Sensor Networks.

Résumé

Résumé :

Ces dernières années, Les réseaux de capteurs sans fil(RCSFs) gagnent en popularité dans la communauté scientifique et industrielle en raison de son déploiement et de ses applications relativement simples. Ces réseaux sont la cible de plusieurs menaces à la sécurité, car les RCSFs fonctionnent habituellement dans un environnement sans surveillance, pour cela les chercheurs mettent en œuvre des solutions de sécurité efficaces qui permettent de protéger le réseau. Dans le contexte, les IDS (Intrusion Detection System) sont une bonne option pour mieux protégé les RCSFs. Les chercheurs comme Almomani et all ont créé un dataset spécialisées pour le RCSFs afin de mieux détecter et classifier quatre types d'attaques par déni de service (Dos) : Blackhole, Grayhole, Flooding et TDMA (Time Division Multiple Access). Ils utilisent un protocole LEACH (Low-Energy Adaptive Clustering Hierarchy.) qui est l'un des protocoles de routage hiérarchisés les plus populaires du RCSFs. De plus, un schéma a été défini pour collecter des données à partir de Network Simulator 2 (NS-2) et ensuite traité pour produire 23 attributs, mais dans ce rapport on à utiliser seulement 19 attributs. Dataset recueillies par les chercheurs est appelé WSN-DS (Wireless Sensor Network-Detection System). Artificial Neural Network (ANN) a été formé sur l'ensemble de données pour détecter et classer différentes attaques DOS. Ce mémoire a mené une expérimentation à l'aide Waikato Environment for Knowledge Analysis (WEKA) pour évaluer l'efficacité des algorithmes d'apprentissage automatique. Tels que K-Plus Proches Voisins, Régression Logistique, Machine vectorielle de soutien, L'arbre de décision, Naïve Bayes.

Mot clé : réseaux de capteurs sans fil, système de détection d'intrusion, attaque DOS, WSN-DS, algorithme d'apprentissage automatique.

Abstract

Abstract:

In recent years, Wireless Sensor Network (WSN) is gaining popularity in the scientific and industrial community due to its relatively simple deployment and applications. These networks are the siphon of several security threats, as WSN typically operates in a hostile, unattended environment, so researchers implement effective security solutions that protect the network. In this context, IDS (Intrusion Detection System) are a good option to better protect the WSN. Researchers such as Almomani and all have created specialized datasets for the WSN to better detect and classify four types of Denial of Service Attacks (DOS): Blackhole, Grayhole, Flooding and TDMA (Time Division Multiple Access). They use a LEACH (Low-Energy Adaptive Clustering Hierarchy) protocol that is one of the most popular hierarchical routing protocols in the WSN. In addition, a schema was defined to collect data from Network Simulator 2 (NS-2) and then processed to produce 23 features, but in this report we use only 19 features. The dataset collected by the researchers is called WSN-DS (Wireless Sensor Network-Detection System). Artificial Neural Network (ANN) has been trained on the dataset to detect and classify different Dos attacks. This report conducted an expression using Waikato Environment for Knowledge Analysis (WEKA) to evaluate the effectiveness of machine learning algorithms. Such as K-Nearest Neighbour (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes.

Keyword: wireless sensor networks, intrusion detection system, DOS attack, WSN-DS, machine leaning algorithms.

ملخص

ملخص :

في السنوات الأخيرة، أصبحت شبكات الاستشعار اللاسلكية (RCSFs: Réseaux de Capteurs Sans Fil) شائعة بشكل متزايد في المجتمع العلمي والصناعي بسبب نشرها وتطبيقاتها البسيطة نسبيًا. هذه الشبكات هي هدف للعديد من التهديدات الأمنية، لأن شبكات الاستشعار اللاسلكية تعمل عادة في بيئة غير مراقبة، لذلك ينفذ الباحثون حلولًا أمنية فعالة تحمي الشبكة. وفي هذا السياق، فإن نظام الكشف عن التسلل (IDS: Intrusion Detection System) هو خيار جيد لتحسين حماية RCSFs. أنشأ باحثون مثل Almomani وجميعهم مجموعات بيانات متخصصة لـ RCSFs لاكتشاف وتصنيف أربعة أنواع من هجمات رفض الخدمة (DOS: Denial Of Service) بشكل أفضل: Blackhole و Grayhole و Flooding و TDMA (Time Division Multiple Access).

يستخدمون بروتوكولا (LEACH: Low-Energy Adaptive Clustering Hierarchy)، وهو أحد أكثر بروتوكولات التوجيه الهرمية شيوعًا في RCSFs. بالإضافة إلى ذلك، تم تعريف المخطط لجمع البيانات من (Network Simulator 2 :NS-2) ثم تمت معالجته لإنتاج 23 سمة، ولكن في هذا البحث تم استخدام 19 سمة فقط. مجموعة البيانات التي جمعها الباحثون تسمى (WSN-DS:Wireless Sensor Network-Detection System) ، تم تدريب الشبكة العصبية الاصطناعية (ANN: Artificial Neural Network) على مجموعة البيانات لاكتشاف وتصنيف هجمات DOS المختلفة. أجرى هذا البحث تجربة باستخدام (WEKA:Waikato Environment for Knowledge Analysis) لتقييم فعالية خوارزميات التعلم الآلي. مثل: K-Plus Proches Voisins و Régression Logistique و Machine vectorielle de soutien و L'arbre de décision و Naive Bayes.

الكلمات المفتاحية: شبكات الاستشعار اللاسلكية، نظام الكشف عن التسلل، هجوم DOS، WSN-DS، خوارزميات التعلم الآلي.

Introduction générale

Introduction Générale :

Les réseaux de capteurs sans fil (RCSFs)sont devenus un domaine de recherche de plus en plus important à cause de leur large gamme d'applications en temps réel la surveillance militaire, la surveillance de la sécurité des bâtiments, la surveillance des feux de forêt et les soins de santé **[1]**. Un RCSFs est composé d'un grand nombre de nœuds de capteurs autonomes, qui sont répartis dans différents domaines d'intérêt pour la collecte de données importantes et transmettre de façon coopérative les données collectées sans fil à un nœud plus puissant appelé «Sink » **[2]**.Cette collecte d'informations est soumises aux contraintes en ressources des capteurs, dont la topologie dynamique, la limitation d'énergie, l'espace de stockage **[3].**

L'économie d'énergie représente l'un des grands défis à soulever pour le bon fonctionnement des réseaux de capteurs. En effet, les nœuds capteurs sont généralement alimentés au moyen d'une petite batterie limitée en puissance, et le replacement de celle-ci est une tâche très difficile voire impossible. Par conséquent, l'épuisement des réserves d'énergie des nœuds capteurs implique la mise hors service du réseau tout entier **[4].** Comme la plupart des réseaux distribués, les RCSFs sont exposés aux menaces de sécurité. En outre, leurs caractéristiques spéciales les rendent très vulnérables aux attaques malicieuses. En effet, les RCSFs sont généralement déployés dans des zones inconnues sans aucune protection physique, ce qui facilite leur capture et compromission. De plus, l'environnement de communication sans fil permet d'écouter et d'espionner le trafic échangé dans le réseau, ce qui ouvre l'horizon pour lancer plusieurs types d'attaques**.** Un attaquant peut compromettre un nœud de capteur, espionner des messages, injecter de faux messages, modifier l'intégrité des données et gaspiller les ressources du réseau. L'attaque par déni de service (Dos) est considérée comme l'une des attaques les plus courantes et dangereuses qui menacent la sécurité des RCSFs. Donc il est nécessaire d'utiliser des mécanismes efficaces pour protéger ce type de réseau contre les attaques Dos.

Toutefois il est bien connu, que les systèmes de détection d'intrusions (IDS) sont des mécanismes de sécurité très efficaces pour protéger le réseau contre les attaques malveillantes, les IDS est nécessaire pour détecter les attaques connues et inconnues et alerter les nœuds de capteurs à leur propos **[5].** Les techniques d'apprentissage automatique considérées comme l'une des principales méthodes pouvant être utilisées avec les IDS pour améliorer leur capacité à identifier et reconnaître les attaquants. La méthode de classification Machine Learning a été utilisée pour détecter les dénis de service dans les RCSFs. **[6]**

❖ Le but de cette mémoire est d'évaluer les techniques d'apprentissage automatique **«K-Plus Proches Voisins, Régression Logistique, Machine vectorielle de soutien, L'arbre de décision, Naïve Bayes.»** pour la détection des attaques DOS dans les RCSFs .Ces techniques sont testées sur un dataset appelé «WSN-DS » contenant des scénarios normaux et des scénarios d'attaque.

❖ Nous avons structuré notre mémoire en trois chapitres :

- **Chapitre 1** : « Les réseaux de capteurs sans fil », nous avons expliqué bien que brièvement la définition et les types de capteurs et leurs fonctionnements, la définition et l'architecture des RCSFs, Cela nous a permis de découvrir les multiples domaines d'utilisation de ces derniers vu la facilité d'utilisation, et aussi nous avons expliqué l'architecture d'un Protocole dans les RCSFs et les contraintes .
- **Chapitre 2** : « La sécurité dans les réseaux capteur sans fil », nous avons parlé sur les objectives de la sécurité dans les RCSFs, les attaques qui ciblent le réseau. Et nous avons parlé sur des mécanismes de sécurité pour répondre aux questions de sécurité dans RCSFs.
- **Chapitre 3** : « l'utilisation l'apprentissage automatique pour détecter les intrusions dans les réseaux de capteurs sans fil », nous avons donné une vue générale sur l'apprentissage automatique telle que les catégories et leurs algorithmes, les mesures d'évaluations, en plus on a donné une brève historique d'un dataset et leurs attributs, après ça on a donné l'évaluation et les résultats obtenus pour chaque algorithmes utilisés.
- Ce travail est terminé par une conclusion générale et des perspectives.

Chapitre 1

Les réseaux de Capteurs sans fil

I Les réseaux de capteurs sans fil

I.1 Introduction :

Depuis leur création, les RCSFs ont connu un succès sans cesse croissant au sein des communautés scientifiques et industrielles. Grâce à ses divers avantages, cette technologie a pu s'instaurer comme acteur incontournable dans les architectures réseaux actuelles. Il offre en effet des propriétés uniques, qui peuvent être résumées en trois points : la facilité du déploiement, diffusion de l'information et le coût réduit d'installation.

En raison de la croissance continue des réseaux de capteurs sans fil, le besoin de mécanismes de sécurité plus efficaces augmente également. Les préoccupations en matière de sécurité du réseau decapteurs devraient être abordés dès le début de la conception du système, car les réseaux de capteurs interagissent avec des données sensibles et fonctionnent habituellement dans un environnement hostilesans surveillance.

I.2 Définition :

I.2.1 Un Capteur :

Le capteur est un instrument de mesure qui permet de transformer une grandeur physique ou chimique observée (température, humidité, l'accélération, les vibrations, etc.) en un signal électrique **[7].**

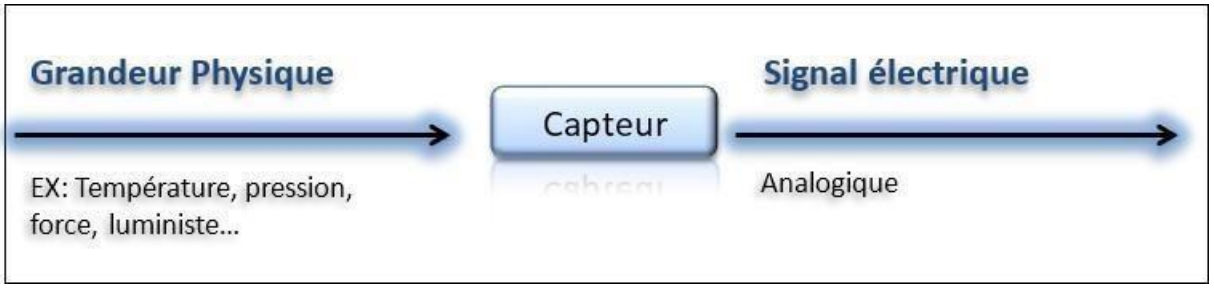

Figure I-1 : Schéma simple d'un capteur **[8]**

Il Ya plusieurs catégories de capteur selon la nature de l'opération. Il est également possible de les classer en capteurs actifs ou passifs**. [9]**

I.2.1.1 Capteur actif :

Les capteurs actifs transforment directement la grandeur physique en grandeur électrique (fournit sa propre énergie), **EX : Thermocouple, capteur de pression etc. [9]**

I.2.1.2 Capteur passif :

Les capteurs passifs nécessitent une source d'énergie électrique extérieure pour obtenir un signal électrique (nécessite une alimentation pour fournir l'information). **EX : capteur de position, humidité etc. [9]**

I.2.2 Réseau de capteurs sans fil :

Un réseau de capteurs sans fil est un cas particulier d'un réseau ad hoc qui constitué d'un grand nombre de nœuds, qui se caractérise par leurs contraintes d'énergie. Qui sont des micro-capteurs capables de collecter et de transmettre des données d'une manière autonome. La position de ces nœuds éparpille aléatoirement dans une zone géographique, nommé « champ de captage » correspondant au terrain concerné pour le phénomène capté. **[10]**

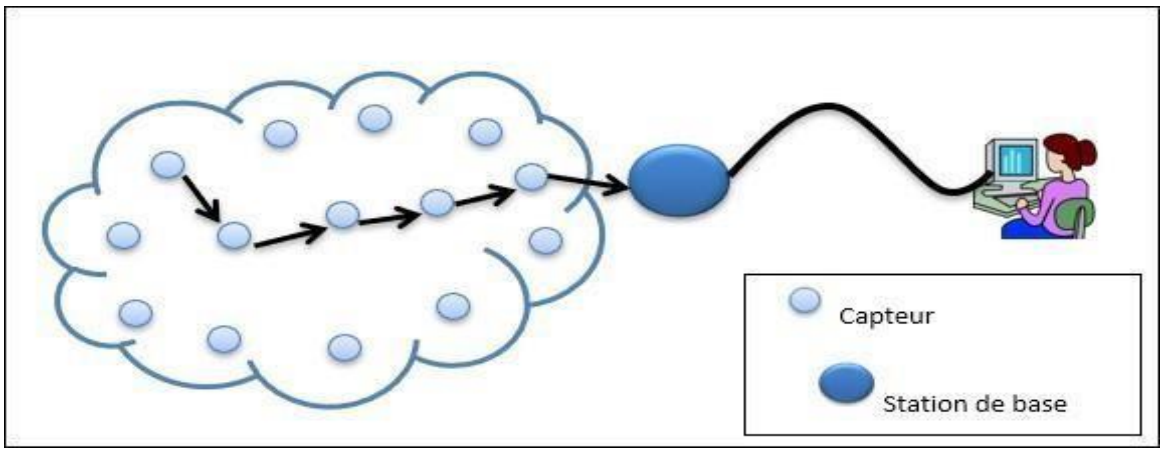

Figure I-2 : schéma simple de réseau de capteur sans fil [10]

I.3 Architecture d'un réseau de capteurs sans fil :

Un réseau de capteurs est constitué essentiellement de plusieurs nœuds :

- **Les nœuds** sont des capteurs: Leur type, leur architecture et leur disposition géographique dépendent de l'exigence de l'application en question. Leur énergie est souvent limitée puisqu'ils sont alimentés par des piles. **[11]**

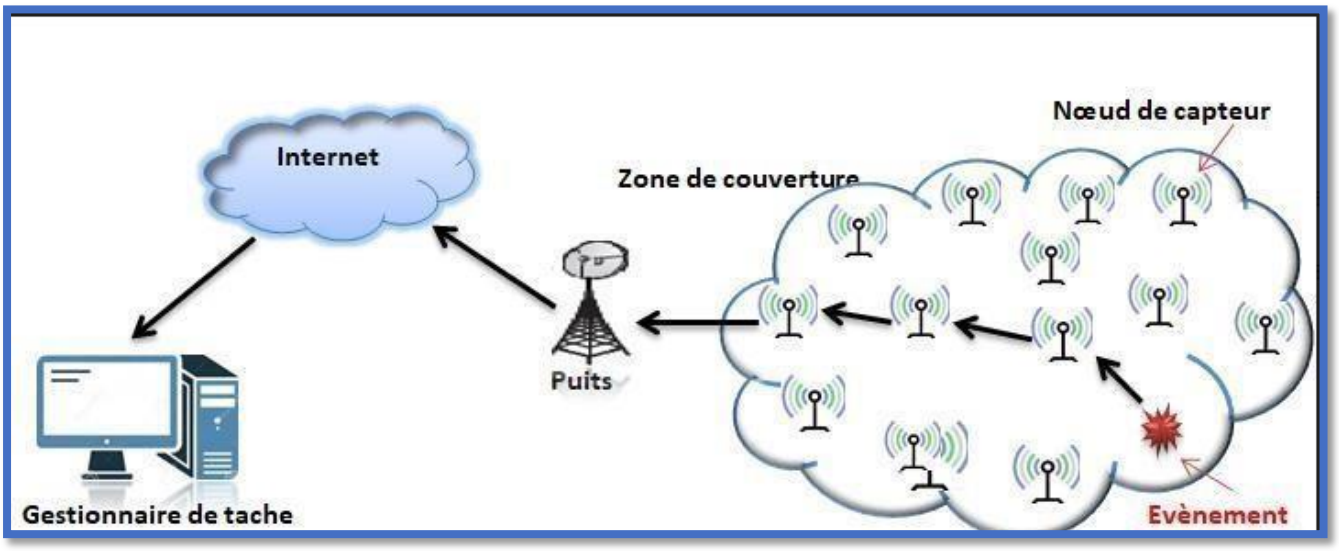

Figure I-3 : Architecture de réseau de capteurs sans fil [10]

Comme le montre la Figure I.3, un RCSFs est composé d'un plusieurs nombre de nœuds capteurs s'éparpille sur le champ de captage. Tous ces nœuds communiquent entre eux, chaque nœud peut communiquer avec les autres nœuds qui sont situés dans sa zone de couverture. Quand la station de base ou encore parfois puits (sink en anglais) diffuse une requête, les nœuds collaborent entre eux pour lui envoyer les informations captées à travers une architecture multi-sauts. Le puits transmet ensuite ces données par Internet ou par satellite au gestionnaire de tâches pour les analyser et prendre des décisions. **[10]**

Il existe deux types d'architectures pour les réseaux de capteurs sans-fil:

I.3.1 Les réseaux de capteurs sans fil plats :

Un réseau de capteurs sans-fil plat est un réseau homogène, ou tous les nœuds sont identiques en termes de fonctionnement, la communication et la complexité du matériel, seul le puit échappe à cette règle puisqu'il joue le rôle d'une passerelle et qui est responsable de la transmission des données issues des différents nœuds capteurs collectées à l'utilisateur final. **[11]**

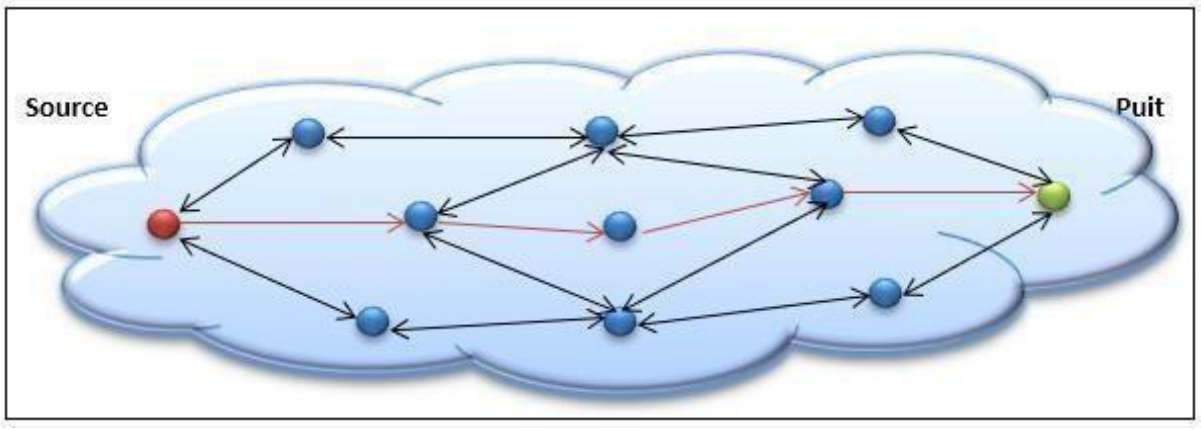

Figure I-4 : Architecture plats [11]

I.3.2 Les réseaux de capteurs sans fil hiérarchiques :

Une architecture hiérarchique a été proposée pour réduire la complexité de la plupart des nœuds capteurs et leur déploiement, en introduisant un ensemble de nœuds capteurs plus puissants. Ceci permet de décharger la majorité des nœuds simples à faible coût de plusieurs fonctions du réseau. **[11]**

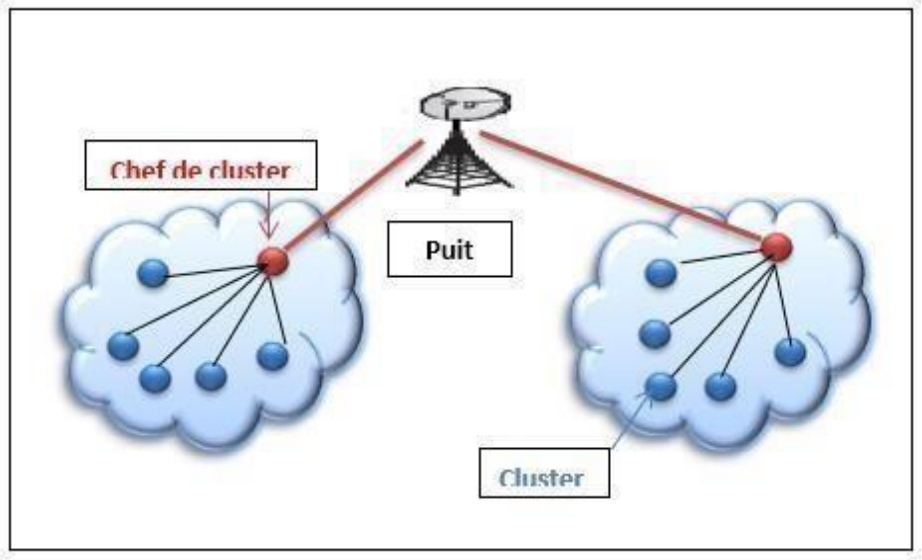

Figure I-5: Architecture hiérarchique [11]

I.4 Architecture d'un nœud de capteur :

Un capteur contient quatre unités de base : **l'unité de capture, l'unité de traitement, l'unité de communication,** et **l'unité d'énergie**. Selon le domaine d'application, des modules supplémentaires peuvent être ajoutés tel **qu'un système de localisation (GPS)**, et **un système mobilisateur**. **[12]**

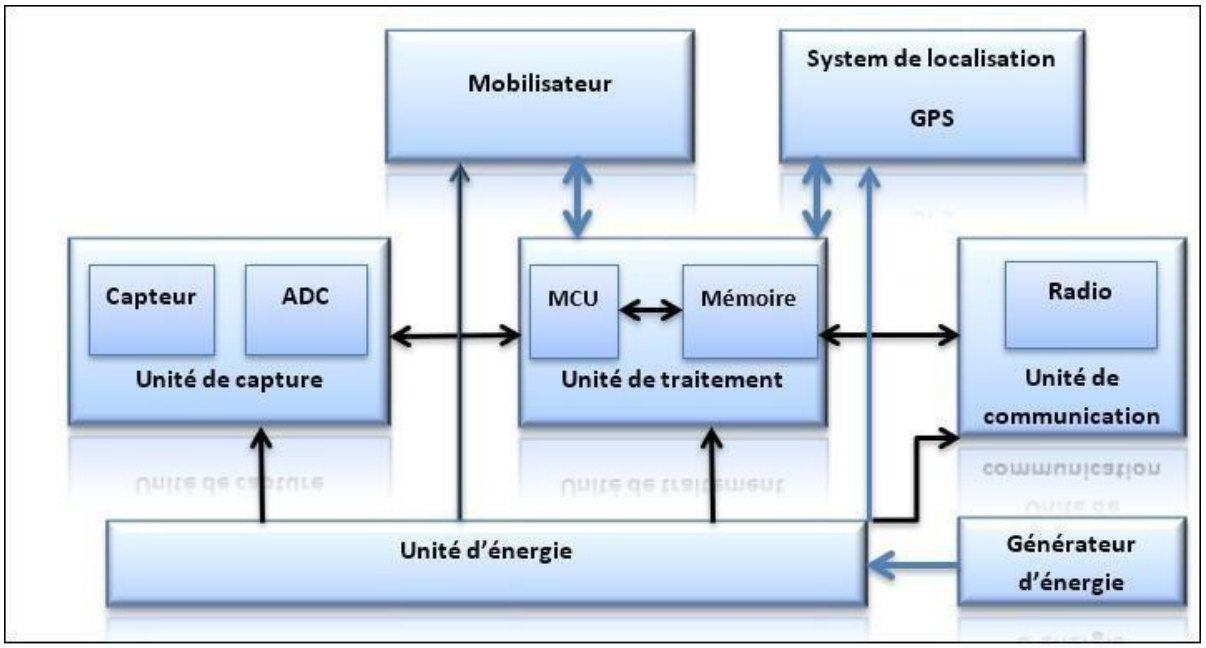

Figure I-6 : Architecture matérielle d'un capteur [12]

❖ **Unité de capture :** C'est l'unité qui est responsable de la capture des grandeurs physiques (Chaleur, Température, pression, force, etc..) et de les transformer en grandeurs numériques (un signal électrique). Plus une unité ADC (Analog to Digital Converters). Le rôle de celui-ci est de transformer le signal analogique produit par les capteurs en un signal numérique compréhensible par l'unité de traitement. **[12]**

❖ **Unité de traitement** : Composé de deux interfaces l'une avec l'unité de capture et l'autre avec l'unité de communication, son rôle consiste à contrôler le bon fonctionnement des autres unités. Cette unité permet l'exécution de procédures de communication qui permettent la coopération d'un nœud avec les autres nœuds du réseau. **[12]**

❖ **Unité de communication :** Cette unité permet de connecter les nœuds du réseau. Un module radio (émetteur/récepteur) est intégré à cette unité qui permet de communiquer entre les divers nœuds du réseau. Elle est responsable de la transmission-réception des données captées et traitées via un canal de communication sans fil. **[12]**

❖ **Unité d'énergie :** Dans le cas des RCSFs, l'unité d'énergie est l'élément le plus important, représentant généralement une batterie, Elle alimente les unités que nous avons citées et elle n'est généralement ni rechargeable ni remplaçable. La faible capacité énergétique au niveau du capteur est la principale contrainte lors de la conception d'un RCSFs**. [12]**

I.5 Les domaines d'applications des réseaux de capteurs Sans fil :

Parmi les domaines d'applications ou les réseaux de capteurs sans fil se révèlent très utiles et leur déploiement à une importance cruciale, on trouve :

- **Application militaire :**

Les capteurs sont déployés de façon autonome afin d'aider les militaires dans ses missions en surveillant et collecte les informations sur la position de l'ennemi ou en analysant le champ de bataille avant d'y envoyer les troupes **[13],** et aussi détection les menaces chimiques, bactériologique **[10].**

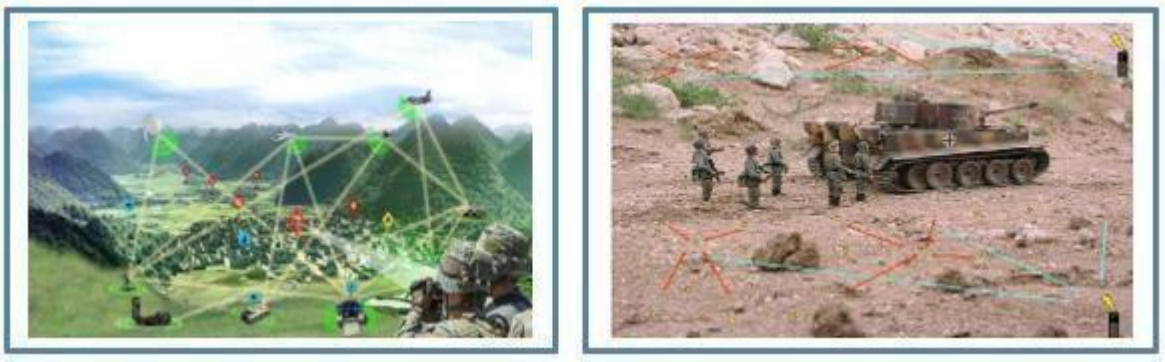

Figure I-7 : Application militaire [14]

- **Application médicale :**

Les RCSFs ont été intégrés au domaine de la médical afin de fournir une surveillance permanente des patients **[13],** Des micro-caméras peuvent être avalées existent déjà, pouvant sans recours à la chirurgie, transmettre des images de l'intérieur d'une fraternité humain. Ils peuvent aussi faciliter la détection d'un cancer, surveiller le rythme cardiaque **[10],** Un capteur portable, il est facile de vérifier le niveau d'oxygéné et le pouls dans le corps humains **[16].**

.

Figure I-8 : Application médicale [15,16]

- **Applications environnementales :**

Ils peuvent aider dans la surveillance météorologique, la détection des dangers naturels, la détection de la pollution (la pollution des eaux, l'air et le sol), faciliter la lutte contre les incendies dans les forêts ou des usines**. [13]**

Figure I-9 : Application environnementales **[17,18]**

- **La domotique :**

Il est possible de charger les capteurs dans des appareils, comme les aspirateurs, les fours à micro-ondes, les réfrigérateurs, les magnétoscopes, etc. Grace à ces micro-capteurs les utilisateurs peuvent plus facilement contrôler ces appareils domestiques localement ou à distance via un réseau externe. Ils peuvent également être déployés pour constituer un système de sécurité en déclenchant une alerte lorsqu'un intrus est présent. **[13]**

Figure I-10 : Application domotique **[19]**

- **Application commerciale :**

Le nœud de capteur peut améliorer le stockage et la livraison et le réseau résultant formé peut être utilisé pour déterminer l'emplacement, l'état et la direction du paquet ou d'une cargaison. Le client qui attend le paquet peut alors obtenir un avis de livraison en temps réel et connaître L'emplacement du paquet. Grâce aux réseaux de capteurs, les entreprises peuvent offrir de meilleursservices tout en réduisant leurs coûts. **[10]**

Figure I-11 : Application commerciale **[20]**

- **Application à la sécurité :**

Les structures d'avions, navires, automobiles, métros, etc. Peut être surveillé en temps réel à travers des réseaux de capteurs ainsi que des réseaux de circulation ou de distribution d'énergie. La surveillance de routes ou voies ferrées pour prévenir des accidents avec des animaux ou des êtres humains ou entre plusieurs véhicules est une des applications envisagées des réseaux de capteurs**. [10]**

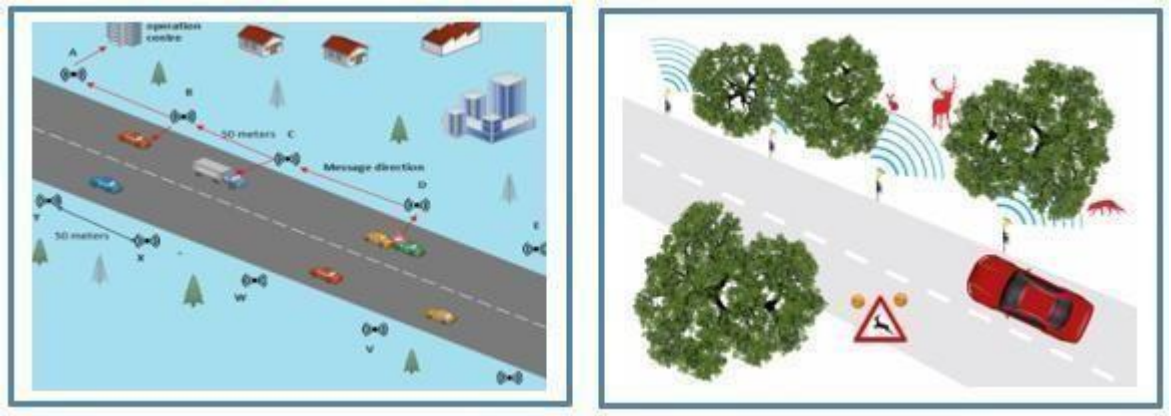

Figure I-12 : Application sécurité

I.6 La Pile protocolaire :

La pile de protocolaire pour les RCSFs se compose de cinq couches de protocole standard. Pour satisfaire les caractéristiques typiques des capteurs. Présentée sur la figure 1.13. Ces couches abordent la dynamique des réseaux et l'efficacité énergétique. **[21]**

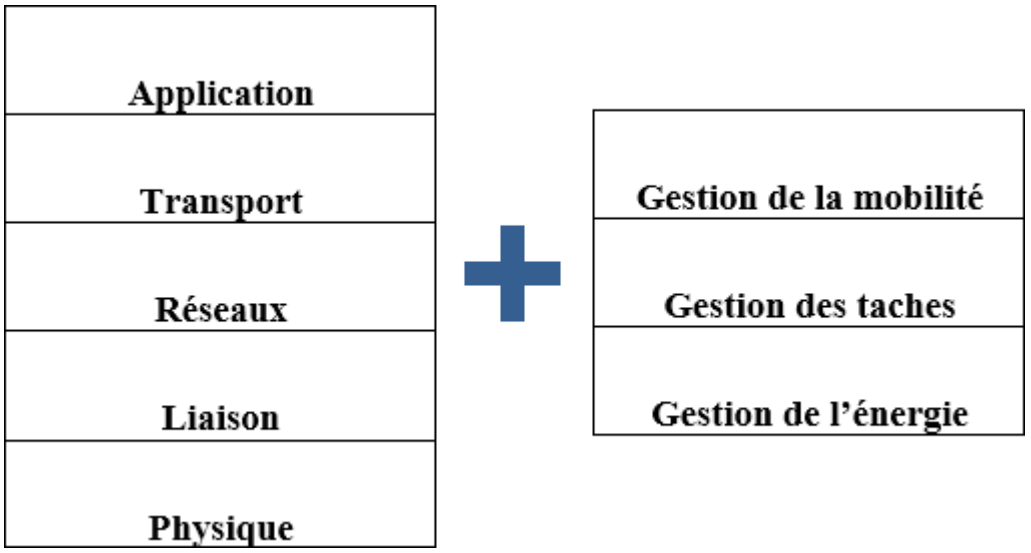

Figure I-13 : Pile protocolaire dans les RCSFs [21]

Couche application : Cette couche fournit l'interface avec l'application. C'est donc le niveau Plus proche de l'utilisateur, directement géré par le logiciel. **[22]**

- ✓ Il existe plusieurs protocoles pour la couche application dont : **[21]**

❖ **Sensors management protocol (SMP)** est un protocole utilisé par l'administrateur pourcommuniquer avec les capteurs en utilisant leurs attributs de nomenclature et leurs localisations.

❖ **Task Assignment and Data Advertisement Protocol (TADAP**) assigne des tâchesaux capteurs pour une meilleure coordination du routage, et de la collecte d'informations.

Couche transport : Cette couche est chargée de transmettre les données, de les diviser en paquets, Contrôle de flux, conservation et gestion éventuelle de l'ordre des paquets Erreur de transmission**. [22]**

- ✓ Parmi les protocoles dans la couche transport citons **: [21]**

❖ **User Datagram Protocol Like (UDP-Like)** ressemble au protocole UDP, mais prenden considération la limite en énergie des capteurs.

Couche réseau : Cette couche permet de trouver une route et une transmission fiable des données, captées, des nœuds capteurs vers le puits en optimisant l'utilisation de l'énergie des capteurs. **[22]**

- ✓ Il existe plusieurs protocoles pour la couche réseau dont **: [21]**

❖ **Low-Energy Adaptive Clustering Hierarchy (LEACH)** permet aux ”Clusters-Head”de collecter des données, de les agréger, puis de les envoyer à la station de base.

❖ **Sequential Assignment Routing (SAR**) permet de trouver un cheminement possible pour envoyer des données selon leurs priorités, selon la consommation de l’´énergie, et selon la qualitéde service.

Couche liaison : Spécifie comment les données sont expédiées entre deux nœuds dans une distance d'un saut. Elle est responsable du multiplexage des données, du contrôle d'erreurs, de l'accès au media ainsi d’assurer la liaison point à point et multi-point dans un réseau de communication**. [22]**

- ✓ Il existe plusieurs protocoles pour la couche liaison dont : **[21]**

❖ **Self-organizing Medium Access Control for Sensors networks (SMACS**) permet auxcapteurs de se construire un réseau de communication sans faire appel à un capteur spécial ”Cluster- Head”. Avec ce protocole, les capteurs communiquent entre eux avec des fréquences sans limites de bande passante. Ces capteurs se mettent au repos lorsqu’ils n’ont pas de données à envoyer.

❖ **Eavesdrop And Register (EAR)** permet d’´établir et de libérer une connexion

De plus, cette pile possède trois niveaux de gestion **[21]:**

❖ **Plan de gestion de mobilité :** permet de détecte et enregistre le mouvement du nœud capteur pendant la phase de routage

❖ **Plan de gestion des tâches :** permet d’assure l’équilibrage et affecter des tâches sur les différents nœuds du réseau afin d’assurer un travail coopératif et efficace en matière de consommation d’énergie.

❖ **Plan de gestion de l’énergie** : permet de gérer et conserver le maximum d’énergieconsommer par les capteurs.

I.7 Contraintes des réseaux de capteurs sans fil :

Les réseaux de capteurs sans spécifiques avec un nombre de nœudsplus conséquents, une énergie limitée et une puissance de calcul plus faible que les réseaux ad-hoc classiques. Ce sont ces particularités que nous introduisons dans la partie suivante. **[23]**

1. La topologie dynamique :

Il est possible que La topologie d'un réseau de capteurs change avec le temps. La dynamité du réseauest causé par des pannes de nœuds à cause d'un endommagement physique ou l'expiration de son énergie. Cela peut causer des cassures de liens entre les nœuds capteurs**.[13]**

2. Capacité de mémoire :

Les capteurs disposent de mémoire vive (RAM) et d'un peu d'espace de stockage, mais ils nesont pas du tout conçus pour sauvegarder de grandes bases de données. Les informations récoltées doivent être acheminées à la station de base, et non stockées sur le long terme par les capteurs eux- mêmes. **[23]**

3. Tolérance aux fautes :

Les capteurs sont des entités sensibles aux altérations d'états comme des phénomènes climatiques (humidité, chaleur, électromagnétisme) ou du fait d'une batterie faible. Dans ce cas de figure un ou plusieurs capteurs peuvent ne pas fonctionner correctement donc le réseau doit être capable de détecter ce type d'erreur est d'y remédier, en cherchant par exemple à modifier ses tables de routage pour trouver un autre chemin permettant de transmettre l'information**. [23]**

4. Mise à l'échelle :

Le nombre de capteurs utilisés dans les réseaux de capteurs sans fil peut varier de quelques entités à plusieurs dizaines de milliers. C'est d'ailleurs la principale utilité des réseaux de capteurs quidoivent pouvoir s'auto organiser à une grande échelle et être efficace quel que soit le nombre. Pour cela les protocoles des réseaux de capteurs sans fil doivent être capables de fonctionner et de s'adapter selonle nombre de nœuds**. [23]**

5. Limitation de l'énergie :

Les capteurs sont équipés de batteries avec une énergie limitée. De plus, les réseaux de capteurs sans fil quand ils sont déployés, le sont souvent dans des zones difficiles d'accès pour l'homme. Il est donc difficile de pouvoir changer les batteries des capteurs. Si le nombre des capteurs dépasse la centaine d'entités, il est encore plus difficile d'intervenir pour trouver le capteur défaillant et changer sa batterie. **[23]**

6. La connectivité :

La connectivité est un problème important dans les réseaux de capteurs. On dit qu'un réseau de capteurs est connecté s'il existe un itinéraire entre chaque paire de nœud. La connectivité dépend essentiellement de l'existence des routes et elle est affectée par les changements de topologie tel que : la mobilité, la défaillance des nœuds, les attaques, etc… ce qui cause la perte des liens de communication, l'isolement des nœuds, le partitionnement du réseau, etc**. [13]**

I.8 Conclusion :

Les réseaux de capteurs sans fil (RCSFs) sont une nouvelle technologie qui a surgit après les grands progrès technologie concernant le développement des capteurs intelligents, des processeurs puissants et des protocoles de communications sans fil. En effet, leurs limites en énergie, en capacité de stockage ainsi que leur déploiement dans des environnements hostiles a fait que ces réseaux sont très vulnérables aux attaques. Le chapitre suivant sera consacré à l'étude de la sécurité dans les réseauxde capteurs sans fil.

Chapitre 2

La sécurité dans les réseaux de capteurs sans fil

II La sécurité dans les réseaux de capteurs sans Fil

II.1 INTRODUCTION :

La sécurité est un domaine très important pour RCSFs, en particulier pour l'application sensibles contre divers types d'attaques DOS.

La sécurité doit interférer avec certaines fonctions sensibles telles que le transfert de paquets, le routage et Gestion du réseau, les fonctions exécutées par certains ou tous les nœuds disponibles dans les RCSFs.

Lorsque nous abordons la question de la sécurité, nous visons à atteindre certains objectifs dont nous parlerons dans ce chapitre, puis nous étudions la classification des attaques qui à leur tour présentent un grand risque contre le réseau, finalement nous parlons des mécanismes de sécurité qui permettront nous à atteindre le niveau de sécurité et de continuité du réseau.

II.2 La sécurité dans les RCSFs:

La sécurité dans RCSFs est un ensemble cohérent de mécanismes, d'algorithmes, de procédures et de schémas permettant d'atteindre et de maintenir un certain niveau de sécurité.

II.3 Objectifs de la sécurité dans le RCSFs :

Lorsque nous parlons sur le problème de sécurité, Cela signifie que nous voulons atteindre un ensemble d'objectifs, dont le plus important :

II.3.1 L'authentification : [24]

L'authentification est le processus effectué par une entité pour vérifier l'identité d'un nœud qui veut communiquer avec d'autres nœuds. Comme les mots de passe, les signatures numériques ou les codes. D'authentification de message.

II.3.2 Autorisation :

L'autorisation est l'opération qui accorder les utilisateurs avec son privilège ce qui était autorisé. Les techniques les plus connus pour imposer l'autorisation sont de maintenir les listes de contrôle d'accès (ACL).

II.3.3 Disponibilité :

Cette propriété sert à garantir que les services réseau sont disponibles même si le réseau de capteur est ciblé par des attaques comme de déni de service et le système inexploitable ou inutilisable.

II.3.4 Intégrité :

Le mécanisme de sécurité doit garantir qu'un message envoyé par un noud capteur à l'autre n'est pas modifié ou altérées par un nœud non autorisé.

II.3.5 Confidentialité :

La confidentialité est un point très important dans la communication des RCSF, elle assure la limitation d'accès à l'information, seules les nœuds autorisées peuvent accéder les données dans le réseau et empêcher de ceux qui sont non autorisé. Les données doivent donc être chiffrées

II.3.6 Contrôle d'accès :

Le service de contrôle d'accès empêche l'utilisation (lecture, écriture, création ou suppression) non autorisée de ressources accessibles par le réseau.

Cette autorisation d'accès a pour but de protéger des personnes, des biens ou des informations. **[25]**

II.3.7 Le non répudiation :

La non-répudiation garantit l'impossibilité de refuser une action ou une transaction en prouvant l'origine des données. En général, ce processus utilise une signature asymétrique en cryptant le hachage du message. **[25]**

Le nœud ne peut pas nier après l'envoi d'un paquet. **[26]**

II.4 Les attaques contre les réseaux de capteurs sans fil :

Une attaque est un ensemble de techniques informatiques ayant pour but de rendre indisponible un service, d'empêcher les utilisateurs légitimes d'un service de l'utiliser.

II.4.1 Classifications des attaquants : [27]

Différents types de modèles d'attaquants avec différentes motivations peuvent mener une même attaque, ce qui rend la modélisation d'un attaquant essentielle dans l'étude de la sécurité des réseaux de capteurs. La modélisation d'un attaquant dépend du type de l'attaque à exécuter, de sa position par rapport au réseau et du nombre d'adversaires utilisés. Dans un réseau de capteurs, un attaquant peut être classifié selon plusieurs critères.

II.4.2 Selon sa position par rapport au réseau :

- Attaquant externe : où l'attaquant est considéré comme un "étranger" par rapport au réseau. Il s'agit d'un utilisateur non autorisé qui s'introduit depuis l'extérieur du périmètre de sécurité du réseau.

➢ Attaquant interne : où l'attaquant se manifeste comme une entité légitime du réseau autorisée à accéder aux ressources fournies par le système. L'attaquant est ainsi authentifié et reconnu par l'ensemble des éléments du réseau

II.4.3 Selon sa capacité : [27]

➢ Attaquant fort : l'attaquant est équipé de ressources supplémentaires par rapport à l'ensemble des nœuds présents dans le réseau. Par exemple, un attaquant utilise un PC portable avec un médium radio sophistiqué.

➢ Attaquant ordinaire : l'attaquant possède les mêmes caractéristiques que les autres nœuds du réseau. De ce fait, il n'a aucun avantage par rapport aux nœuds légitimes.

II.4.4 Selon son nature :

Du fait qu'un nœud malicieux peut opérer soit au niveau des données échangées entre les nœuds ou au niveau de la topologie du réseau, ses attaques peuvent être classées en deux catégories : actives et passives.

II.4.4.1 Attaquant passif : [28]

Dans ce type d'attaques, les attaquants sont typiquement camouflés (cachés), le rôle de ces attaques se centralise dans un prélèvement ou une écoute de l'information sur le réseau, ce qui permet à l'attaquant contrôler et d'observer les données entre les nœuds.

Ce type d'attaques est facile à réaliser et en même temps difficiles à détecter. **[28]**

Les attaques de cette catégorie peuvent être regroupées dans les types suivants :

- **Camouflage d'adversaires : [28,29]**

Suite à une attaque active, un adversaire peut être compromis ou inséré dans un chemin de routage comme étant un nœud légitime pour attirer des paquets afin d'analyser le trafic dans une région.

- **Écoute (Eavesdropping) :**

L'écoute passive est définie comme l'acte d'écouter discrètement une conversation privée. Un attaquant s'insère dans un chemin actif, afin d'écouter passivement tout le trafic envoyé sur le support de diffusion et extraire les données collectées par l'ensemble du réseau (données agrégées). Si aucun mécanisme cryptographique n'est utilisé pour protéger les messages, l'adversaire pourrait facilement comprendre le contenu de la conversation, ce qui menace la confidentialité des données.

- **Analyse du trafic : [29]**

En raison de l'analyse approfondie du trafic, un adversaire combine l'écoute et l'analyse de trafic pour réaliser une attaque efficace. Par conséquent, il est possible d'obtenir des informations utiles sur la topologie de réseau, comprendre les rôles des nœuds ou d'identifier la station de base.

Par exemple, les contacts d'un nœud peuvent être déterminés en filtrant le trafic du réseau.

II.4.4.2 Attaquant actif : [28,29]

Les attaques actives sont les attaques dans lesquelles l'attaquant tente de modifier l'information ou crée un faux message. **[29],** ils n'affectent pas seulement sur la confidentialité des données, mais peuvent également affecter sur la disponibilité, l'actualité et l'intégrité des données.

Par rapport aux attaques passives, ce type d'attaque peut être détecté en développant des mécanismes de sécurité avancés. **[28]** On regroupe les attaques actives les plus connues selon les classes suivantes :

1. Attaques de la couche physique

Cette couche responsable de la spécification des câbles, de la fréquence porteuse…etc. Elle doit fournir des techniques de transmission, de réception et de modulation pour fournir des données de haute qualité. **[30]**

Deux attaques sont explorées dans cette couche :

- **Attaque de brouillage (Jamming) :**

C'est une attaque de type Déni de Service (DoS) dont le but est de perturber la communication, vu la sensibilité du canal sans fil, un attaquant utilise un puissant dispositif de brouillage pour interférer le canal de communication entre deux nœuds communiqués. Cette attaque très dangereuse si elle cible la station de base afin de paralyser l'ensemble du réseau ou le chef de cluster pour isoler toute une région.

Afin de défendre les RCSFs contre le brouillage de communication, les techniques de transmission de signaux en commutant rapidement une porteuse parmi de nombreux canaux de fréquence sont employées, ce qui empêche l'attaquant de détecter le canal de fréquence utilisé entre l'émetteur et le récepteur. **[28]**

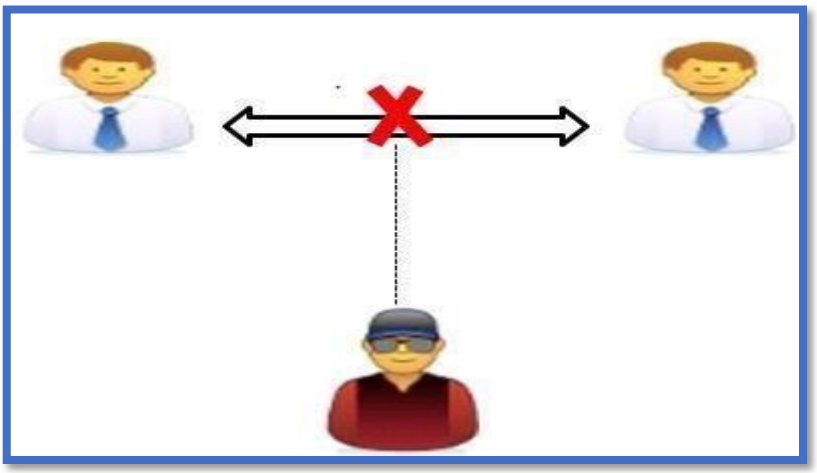

Figure II-1 attaque jamming

- **Attaque d'altération (Tampering) :**

L'altération physique est une autre façon d'attaquer.

Elle consiste à la capture et à l'accès physique au nœud pour exporter toutes les informations importantes comme les clés utilisées pour le chiffrement, altérer les circuits électroniques, reprogrammer le nœud. **[28]**

2. Les attaques de la couche liaison

Elle spécifie comment transférer les données entre deux nœuds dans une distance d'un saut. Responsable du multiplexage des données, du contrôle d'erreurs, de l'accès aux media… etc.

- **Collisions :**

Une collision de données est le résultat de la transmission simultanée de paquets de données entre deux ou plusieurs appareils ou nœuds de domaine de réseau sur la même fréquence. Les paquets de collision de données sont fragmentés et retransmis. **[31]**

Un attaquant peut provoquer de manière stratégique des collisions dans des paquets spécifiques, Les résultats possibles de telles collisions l'épuisement des ressources, l'injustice dans l'allocation, risque d'épuiser la batterie.

Par conséquent, le fonctionnent des applications en temps réel, qui s'exécutent sur d'autres nœuds se dégrade à cause de la perturbation par l'interruption de leurs transmissions de trames. **[28]**

Difficile à détecter tell que la seule preuve d'attaque est les paquets incorrects.

- **Épuisement :**

Cette attaque contrôle les ressources d'alimentation des nœuds en les forçant à retransmettre des messages même en l'absence de collisions. Cela inclut consommer toute la puissance de calcul du processeur, occuper tout l'espace disque ou mémoire du système informatique pour le rendre inutilisable. **[32]**

3. Les attaques au niveau de la couche réseau

Cette couche permet de gérer l'envoi et le routage des données. Les protocoles les plus importants utilisés au niveau de cette couche sont l'IP et l'ICMP. **[33]**

- **Attaque de trou noir (Blackhole) :**

L'attaque par trou noir consiste d'abord à insérer des nœuds malveillants en divers moyens dans le réseau. Le nœud modifiera la table de routage pour appliquer Le nombre maximal de nœuds adjacents par lesquels les informations sont transmises. Alors tous Les informations transitant par ce

nœud.

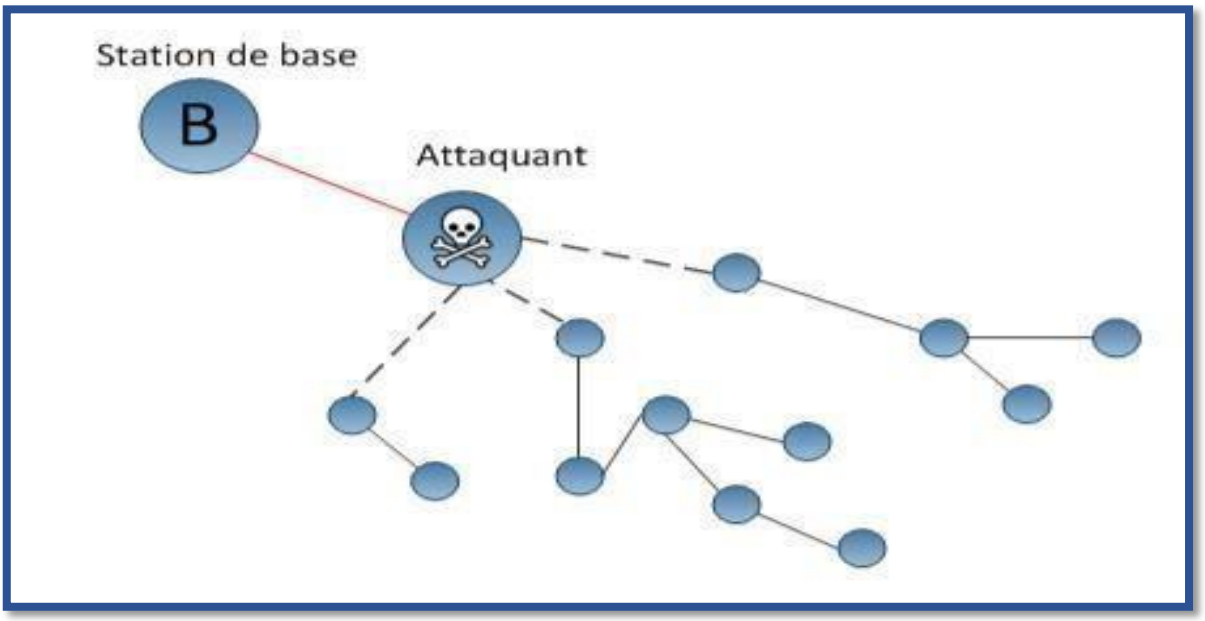

Figure II-2: Attaque de trou noir (Blackhole)

- **L'attaque Sybil (Sybil Attack) :**

Le but de cette attaque est de faire passer un nœud malveillant à travers plusieurs nœuds en supposant plusieurs identités pour créer plusieurs chemins à travers le nœud, alors qu'en fait il n'y a qu'un seul chemin.

Les attaques Sybil visent à assurer la fiabilité du réseau en se basant sur des protocoles qui établissent la redondance des chemins.

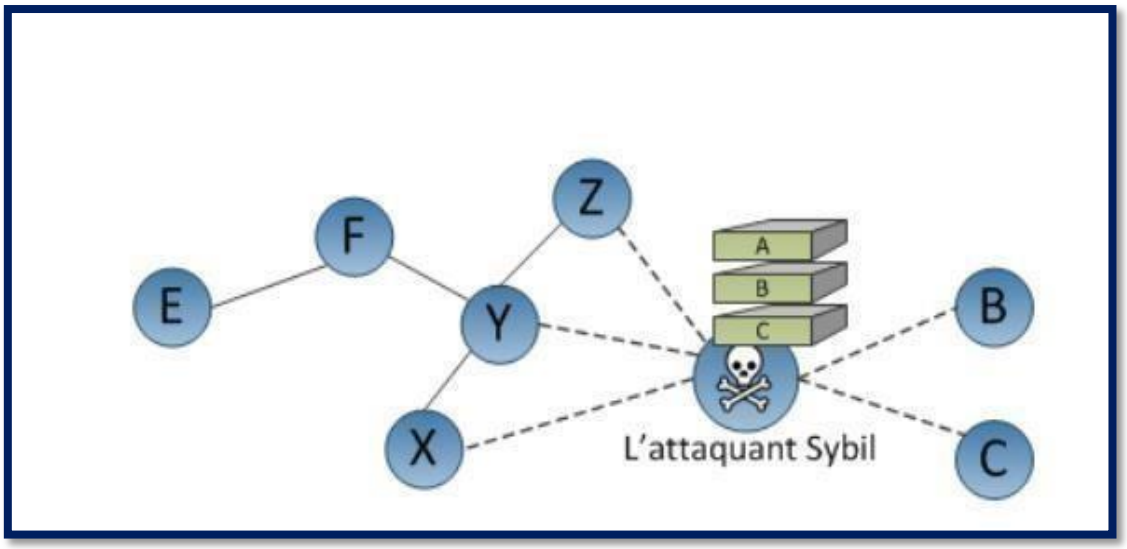

Figure II-3 : Attaquant Sybil

- **L'attaque de trou de puits (Sinkhole Attack) :**

L'attaquant devient attractif en proposant aux nœuds des chemins le plus rapide pour atteindre la station de base avec des connexions puissantes ce qui pousse les nœuds émetteurs à modifier leurs

tables de routage pour acheminer les données par ce nœud malicieux.

Ainsi l'ensemble de ces nœuds va s'adresser en particulier à ce nœud malicieux pour transmettre l'information à la base. Toutes les informations qui transitent vers la base pourront être récupérées par l'attaquant.

- **Attaque du trou gris "Greyhole" :**

Une variante de l'attaque précédente est appelée trou gris, c'est une variante améliorée de l'attaque du trou noir dans laquelle seuls certains types de paquets sont ignorés par le nœud malicieux. Par exemple trou gris va ignorer toutes les informations concernant le routage, sauf pour des informations plus importance.

Ce type d'attaque est ainsi plus difficile à détecter tant qu'il se comporte de manière normale. **[34]**

- **L'attaque de trou de ver (Wormhole Attack) :**

L'attaque du trou de ver nécessite l'insertion d'au moins deux nœuds malicieux Ces deux nœuds sont reliés entre eux par une connexion puissante comme par exemple une liaison filaire.

Ce chemin attaquant-attaquant est un saut et donc plus rapide et plus optimisé, ce qui permet de créer deux nœuds d'attaque de type Sinkhole attack et permet à chacun des deux nœuds de récupérer des informations d'un point du réseau, de les modifier, puis de les transmettre à un autre. Point dans le réseau en ignorant les nœuds intermédiaires. Les informations divulguées seront envoyées à BS. Le protocole pour devenir victime d'une telle attaque est basé sur :

1- la latence de routes

2- la première route découverte

3-le nombre de sauts pour atteindre la destination.

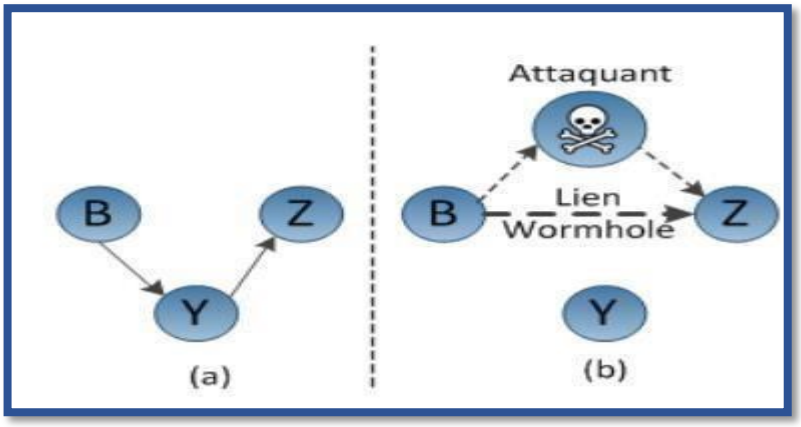

Figure II-4:Attaque Wormhole

- **L'attaque d'inondation par paquet Hello (Hello Flood Attack) :**

L'objectif de cette attaque est de consommer l'énergie des nœuds capteurs, par un envoi continu, à un signal puissant des messages de découverte du voisinage de type HELLO.

Les nœuds destinataires du message essayent de répondre au nœud malicieux. A force de tenter de lui répondre, tous les nœuds concernés par ce message HELLO vont petit à petit consommer l'intégralité de leur énergie.

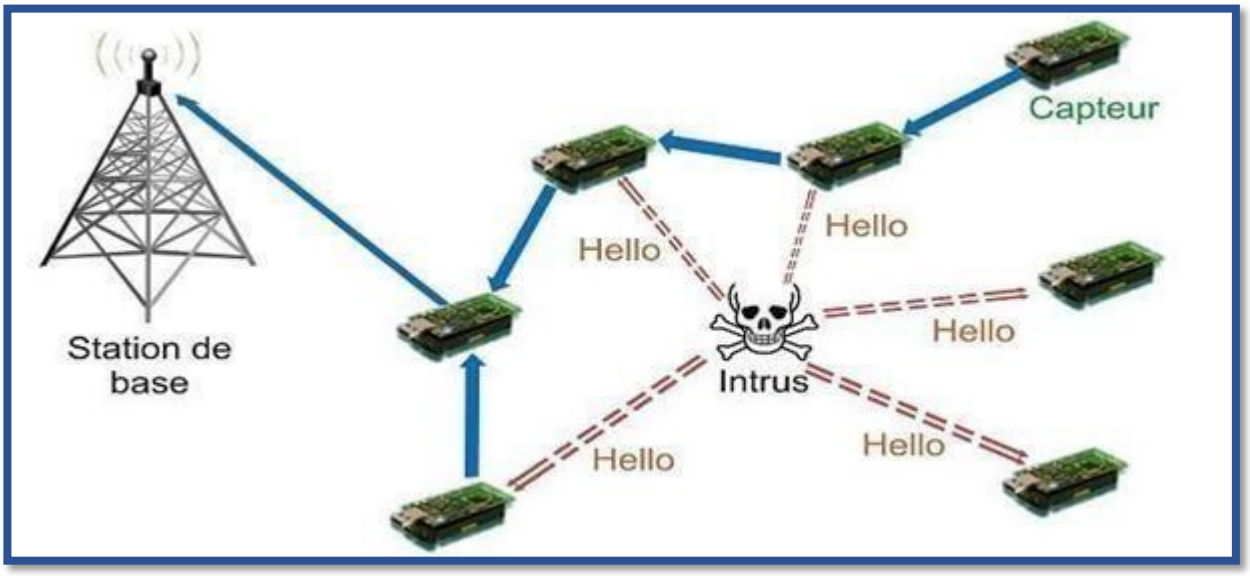

Figure II-5: l'attaque Hello

- **L'attaque par ordonnancement (Scheduling Attack) :**

L'attaque par ordonnancement a été introduite dans une étude précédente des auteurs. L'attaque de planification se produit pendant la phase de configuration du protocole LEACH, lorsque les CH configurent les horaires TDMA pour les intervalles de temps de transmission de données. L'attaquant qui agit en tant que CH attribuera à tous les nœuds le même créneau horaire pour envoyer des données. Cela se fait en changeant le comportement de la diffusion à la planification TDMA unicast. Un tel changement entraînera une collision de paquets qui entraînera une perte de données.

4. Les attaques au niveau de la couche transport :

Cette couche est chargée du transport des données et leur décomposition en paquets et contrôles la qualité du service.

- **Inondation (Flooding) :**

L'objectif de cette attaque est de provoquer un déni de service (DoS). En effet, un ou plusieurs nœuds malicieux du réseau effectuent des envois réguliers de messages à une puissance d'émission forte dans le but de saturer le réseau.

- **Désynchronisation (De-synchronization) :**

Un attaquant tente de perturber une connexion active entre deux nœuds en envoyant de faux messages qui ont changé de numéros de séquence ou de drapeaux de contrôle, en raison d'une désynchronisation, les nœuds concernés renvoient les faux paquets, ce qui gaspille énormément leur énergie.

5. Les attaques au niveau de la couche application :

Cette couche gérer les données générées et utilisables par les applications.

Un attaquant pourrait tenter de submerger les nœuds du réseau avec des stimuli de capteurs, ce qui amènerait le réseau à transférer de gros volumes de trafic vers une station de base. Une autre attaque de la couche application consiste à injecter des paquets parasites ou rejoués dans le réseau.

Les objectifs des attaques contre cette couche sont de consommer la bande passante du réseau et drainer l'énergie des nœuds.

II.5 Les mécanismes de sécurité dans les RCSFs :

Plusieurs mécanismes, basés généralement sur la notion de cryptographie, sont mis en place afin de répondre à la question de la sécurité dans les RCSF.

II.5.1 La cryptographie : [35]

Elle est définie comme étant une science permettant de convertir des informations "en clair" en informations cryptées, c'est à dire non-compréhensibles, et puis, à partir de ces informations cryptées, de restituer les informations originales.

Dans la cryptographie moderne, l'habilité de maintenir un message crypté secret, repose non pas sur les algorithmes, mais sur une information secrète dite clé qui est un paramètre utilisé en entrée d'une opération cryptographique et qui doit être utilisée avec les algorithmes pour produire le message crypté. Il existe plusieurs outils cryptographiques dont nous citons **:**

II.5.1.1 Le chiffrement :

Le chiffrement est le système cryptographique assurant la confidentialité. Pour cela, il utilise des clés. Selon cette utilisation, on distingue deux modes de chiffrement :

- **Le chiffrement symétrique :**

Dans le chiffrement symétrique, une même clé est partagée entre l'émetteur et le récepteur. Elle est utilisée par l'émetteur pour chiffrer le message et par le récepteur pour le déchiffrer.

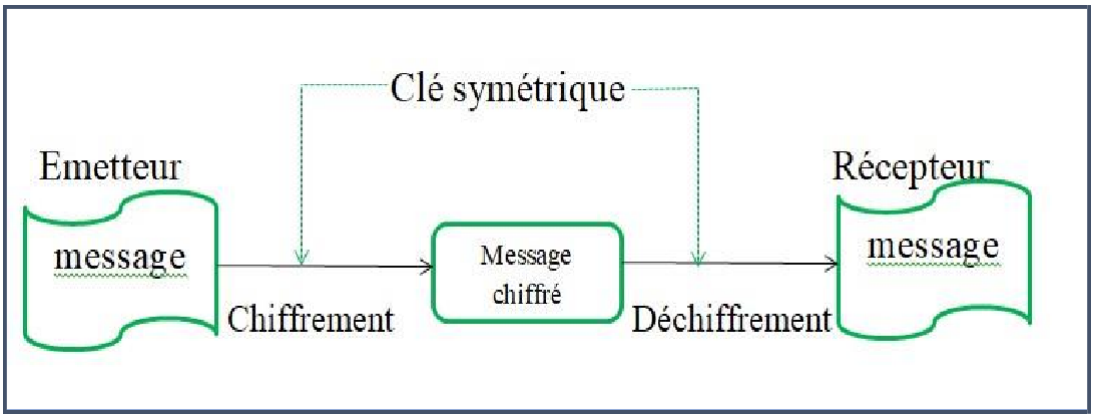

Figure II-6: Le chiffrement symétrique

- **Le chiffrement asymétrique :**

Également appelée cryptographie à clé publique, a été introduite en 1976 par Diffue et Hellman. Le récepteur génère une paire de clés asymétriques, une clé publique qui est diffusée à tous les expéditeurs et une clé privée qui est gardée secrète chez lui. Tout message chiffré avec la clé publique ne peut être déchiffré qu'avec la clé privée correspondante.

Bien que la cryptographie à clé publique présente certains avantages par rapport à la cryptographie à clé symétrique, son utilisation dans RCSF est très coûteuse en termes de ressources, de sorte que la plupart des algorithmes proposés utilisent des méthodes de chiffrement symétriques.

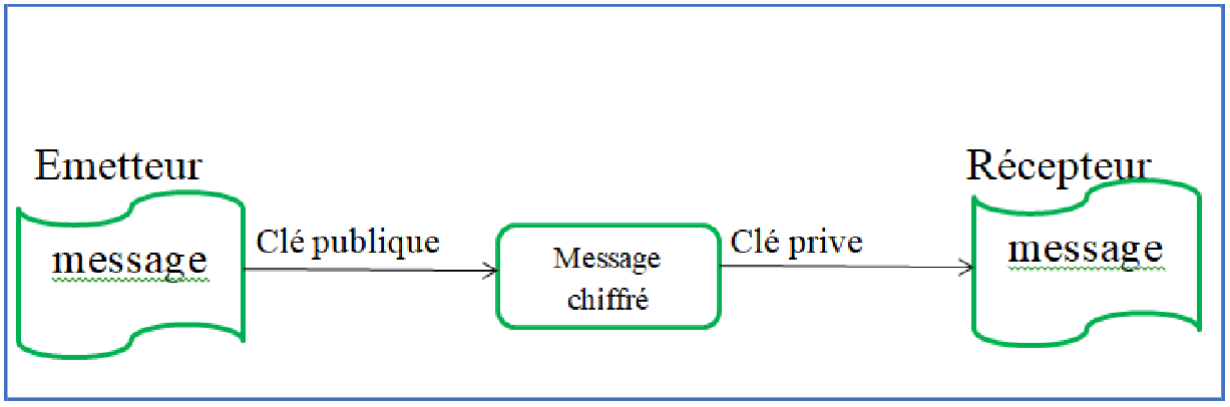

Figure II-7: chiffrement asymétrique

II.5.1.2 La signature digitale :

C'est un système cryptographique assurant la non-répudiation de la source. Elle repose sur les clés asymétriques. L'émetteur signe les données à transmettre avec sa clé privé (A) en produisant une signature digitale. Cette dernière est par la suite envoyée avec les données. Si elle peut être déchiffrée avec la clé publique (A) par le récepteur et si son résultat est identique aux données reçus alors la signature est valide, c'est-à-dire, les données proviennent de leur émetteur légitime qui ne pourra pas nier l'émission de ces données dans le futur.

II.5.1.3 La fonction de hachage :

C'est le mécanisme qui assure l'intégrité de données, Elle permet de générer une chaîne de taille inférieure et généralement fixe à partir d'une chaîne de longueur quelconque. Par conséquent, la chaîne résultante est appelée empreinte. D'un autre cotée, une fonction de hachage est une fonction à sens unique.

L'émetteur utilise la fonction de hachage pour créer une empreinte du message transmettre, puis il transmet le message et l'empreinte vers le récepteur. A la réception du message, le récepteur calcule l'empreinte du message reçu et il la compare à l'empreinte initiale. Si les deux empreintes correspondent. Autrement dit qu'il est facile à calculer l'empreinte d'une chaîne donnée, mais il est impossible de déduire à la chaîne initiale à partir d'une empreinte donnée. **[36]**

II.5.2 Système de détection d'intrusion :

On appelle **IDS** (Intrusion Detection System) un mécanisme écoutant le trafic réseau de manière furtive afin de repérer des activités anormales ou suspectes et permettant ainsi d'avoir une action de prévention sur les risques d'intrusion. [37]

Il existe trois niveaux d'IDS : [38]

- ✓ Les NIDS (Network based Intrusion Detection System) : ils assurent la sécurité au niveau du réseau.
- ✓ Les HIDS (Host based Intrusion Detection System) : ils assurent la sécurité au niveau des hôtes.
- ✓ Les IDS hybrides : qui utilisent les NIDS et HIDS pour avoir des alertes plus pertinentes.

Les IDS disposent de deux approches différentes, afin de déceler les intrusions :

Les IDS à signature : Cette approche consiste à rechercher dans l'activité de l'élément surveillé les empreintes (ou signatures) d'attaques connues. Ce type d'IDS est purement réactif ; il ne peut détecter que les attaques dont il possède la signature.

Une signature permet de définir les caractéristiques d'une attaque, au niveau des paquets (jusqu'à TCP ou UDP) ou au niveau des protocoles (HTTP, FTP...).

Les IDS à anomalie : Elle consiste à détecter des anomalies par rapport à un profil "de trafic habituel". La mise en œuvre comprend toujours une phase d'apprentissage au cours de laquelle les IDS vont découvrir le fonctionnement normal des éléments surveillés. Ils sont ainsi en mesure de signaler les divergences par rapport au fonctionnement de référence.

Ils présentent l'avantage de détecter des nouveaux types d'attaques. Cependant, de fréquents ajustements sont nécessaires afin de faire évoluer le modèle de référence de sorte qu'il reflète l'activité normale des utilisateurs et réduire le nombre de fausses alertes générées.

Les IDS dans le domaine des RCSFs sont devenus assez matures et offrent de bons résultats tant dans la pertinence que dans la fiabilité.

II.6 Conclusion

Dans ce chapitre, nous présentons l'objective de la sécurité dans les RCSF et discutons de diverses attaques ciblant les réseaux.

D'autre part, nous discutons des mécanismes de sécurité, souvent basés sur des concepts cryptographiques, pour répondre aux questions de sécurité dans RCSF. Dans le chapitre suivant, nous sommes intéressés par l'apprentissage automatique pour la détection des attaques Dos dans les réseaux de capteur dans fil.

Chapitre III

Utilisationde l'apprentissage Automatique pour détecter les intrusions dans les réseaux de capteurs sans fil

III Utilisation de l'apprentissage automatique pour détecter les intrusions dans les réseaux de capteurs sans fil.

III.1 Introduction :

La nature et la généralisation de l'utilisation des RCSFs introduisent de nombreuses menaces et attaques de sécurité. Un système efficace de détection des intrusions (IDS) devrait être utilisé pour détecter les attaques. Détecter des telles attaques est difficile, en particulier la détection des attaques par déni de service (DOS). Les techniques de classification par apprentissage automatique ont été utilisées comme approche pour la détection des DOS. Ce chapitre a mené une étude à l'aide de WaikatoEnvironment for Knowledge Analysis (WEKA) afin d'évaluer l'efficacité des algorithmes d'apprentissage automatique pour détecter des attaques Dos, L'évaluation est fondée sur un ensemble de données appelé WSN-DS **[39]** qu'on va détailler dans la suite de ce chapitre.

III.2 L'apprentissage automatique :

L'apprentissage automatique (en anglais machine learning) est utilisé en intelligence artificielle et dans l'analyse des données (Analytics and Data Science) qui se fonde sur des approches mathématiques et statistiques pour donner une capacité d' « apprentissage » à partir de données, c'est à-dire d'améliorer les capacités des machines à résoudre des tâches sans être explicitement programmés.

- Il excite deux catégories de l'apprentissage automatique :

III.2.1 Apprentissage supervisé :

L'apprentissage supervisé commence généralement par un ensemble de données bien défini et une certaine compréhension de la façon dont ces données sont classifiées. L'apprentissage supervisé a pour but de déceler des modèles au sein des données et de les appliquer à un processus analytique. Ces données comportent des caractéristiques associées à des libellés qui définissent le sens des données.

Les modèles de formation supervisée ont une large applicabilité à une variété de problèmes commerciaux, y compris la détection des fraudes, les solutions de recommandation, la reconnaissance vocale ou l'analyse des risques **[40]**

L'apprentissage supervisé peut être utilisé pour deux types de problèmes qui sont:

III.2.1.1 Classification :

La classification est le processus de recherche ou de découverte d'un modèle ou d'une fonction qui aide à séparer les données en plusieurs classes catégorielles, c'est-à-dire des valeurs discrètes. Dans la classification, les données sont classées sous différentes étiquettes en fonction de certains paramètres donnés en entrée, puis les étiquettes sont prédites pour les données nouvelles.

III.2.1.2 Régression :

La régression est le processus de recherche d'un modèle ou d'une fonction permettant dedistinguer les données en valeurs réelles continues au lieu d'utiliser des classes ou des valeurs discrètes.

La régression permet d'utiliser un algorithme dans l'application d'apprentissage automatique pour classer les données entrantes en fonction des données historiques.

III.2.2 Apprentissage non supervisé :

Avec cet apprentissage, vous avez toujours des features, mais pas d'étiquettes (label) car nous nous n'essayons pas de prédire quoi que ce soit. A partir des données historiques dont nous disposons, nous essayons de voir ce que nous pouvons apprendre des données, sans oublier de vérifier les conclusions tirées avec des experts du domaine.

Ce type d'apprentissage automatique est souvent utilisé pour découvrir des structures et des modèles dans les données. Il peut également être utilisé pour l'ingénierie des fonctionnalités dans la préparation des données pour l'apprentissage supervisé.

L'apprentissage non supervisé peut être utilisé pour deux approches : {le regroupement **(clustering)** et **l'association**.}

III.3 Les algorithmes d'apprentissage supervisé :

III.3.1 K-Plus Proches Voisins (K-PPV) :

(K- Nearest Neighbor : K-NN) est un algorithme de raisonnement qui prendre des décisions en recherchant un ou plusieurs cas similaires déjà résolus. La décision consiste à chercher les k échantillons les plus proches de l'objet et de l'affecter à la classe qui est la plus représentative dans ces k échantillons. Cet algorithme peut être utilisé pour résoudre des problèmes de classification et de régression **[41]**

III.3.2 Bayes Naïve :

Naïve Bayes est un algorithme d'apprentissage supervisé, basé sur le théorème de Bayes et utilisé pour résoudre les problèmes de classification.

Bayes naïf est un classificateur probabiliste, ce qui signifie qu'il prédit sur la base de la probabilité d'un objet, il est l'un des algorithmes de classification les plus simples et les plus efficaces qui aide à construire les modèles d'apprentissage automatique rapide pour faire des prédictions rapides**. [42]**

III.3.3 L'arbre de décision :

(Decision Trees : DT) Il s'agit d'une représentation graphique pour obtenir toutes les solutions possibles à un problème de classification et de régression. Il s'agit d'un classificateur arborescent, où les nœuds internes représentent les caractéristiques d'un ensemble de données, les branches représentent les règles de décision et chaque nœud de feuille représente le résultat. **[43]**

III.3.4 Machine vectorielle de soutien :

(Support Vector Machine : SVM) est l'un des algorithmes d'apprentissage supervisé les plus populaires, il est principalement utilisé pour les problèmes de classification ou de régression dans l'apprentissage automatique.

Le but de l'algorithme est de créer la meilleure ligne de décision qui peut séparer l'espace n- dimensionnel en classes afin que nous puissions facilement mettre le nouveau point de données dans la catégorie correcte à l'avenir. Cette limite de meilleure décision est appelée un hyperplan**. [44]**

III.3.5 Régression Logistique :

(Logistique Regression :LR) est l'un des algorithmes d'apprentissage supervisé les plus populaires, il mesure la relation entre la variable dépendante catégorielle et une ou plusieurs variables indépendantes en donnant une estimation à la probabilité d'occurrence d'un événement à travers l'usage de sa fonction logistique.**[45]**

III.4 Les algorithmes d'apprentissage non supervisé :

III.4.1 K-Means :

Le premier algorithme d'apprentissage non supervisé est l'algorithme de classification K-means. Il utilise un raffinement itératif pour produire un résultat final. Les entrées de l'algorithme sont le nombre de clusters et le dataset non étiqueté. L'ensemble de données est un ensemble de caractéristiques pour chaque point de données. Les algorithmes commencent par les estimations

initiales pour les K centrioles. Ces centrioles peuvent être générés de manière aléatoire ou directement à partir du dataset.

III.4.2 Association Rules :

Le système d'association permet de trier et regrouper les données qui peuvent être liées grâce à certaines caractéristiques. Le but est donc de trouver des objets liés les uns aux autres sans qu'il s'agisse néanmoins d'objets identiques.

III.5 WEKA :

III.5.1 Définition :

WEKA «Waikato Environment for Knowledge Analysis» est une suite de logiciels open source permettant d'explorer et d'analyser des fichiers de données développé en Java à l'université de Waikato en Nouvelle-Zélande et publié sous licence GNU General Public License.

Elle contient de nombreux algorithmes pour le regroupement, la classification, les règles d'association et la visualisation des données.

Figure III-1: Weka_(software)_logo

Les avantages de l'utilisation de Weka sont :

- Gratuit.
- Portable.
- Facile à utiliser.
- Large collection de modèles de machine Learning.

III.5.2 Les métriques d'évaluation :

L'évaluation de l'efficacité des modèles de classification utilisés dans la détection DOS est faite avec des métriques de base qui sont : **[46]**

- **True Positive (TP) :** le nombre de cas d'attaques classés correctement comme attaques.
- **False Positive (FP) :** le nombre de cas normaux classés incorrectement comme attaques
- **True Negative (TN) :** le nombre de cas normaux classés correctement comme normaux.
- **False Negative (FN) :** le nombre de cas d'attaques classés incorrectement comme normaux.

A partir de ces métriques nous pouvons calculer les métriques de confusion suivant :

- **Precision :** est le nombre d'éléments qui ont été classés correctement dans chaque classe. Le plus grand nombre de prédictions correctes signifie plus la performance du classificateur. Precision peut être donnée en utilisant l'équation suivante **: [47]**

  Precision= (TP)/ (TP+FP)

- **Recall :** est le pourcentage d'éléments qui ont été classés correctement dans chaque classe comme positifs. Recall peut être effectué à l'aide de l'équation suivante : **[47]**

  Recall= (TP)/ (TP+FN)

- **Accuracy :** Dans les problèmes de classification, représente le pourcentage d'instances correctement classées. C'est une bonne mesure lorsque les classes de variables cibles sont presque équilibrées. Accuracy peut être donnée en utilisant l'équation suivante : **[47]**

  Accuracy= (TP+TN)/ (TP+TN+FP+FN)

III.6 Description « Data Set »:

Almomani et all **[48].** Ont construit un dataset pour les RCSFs, contenant quatre types d'attaques Dos plus un cas normal (non attaque). L'ensemble de données recueillies est appelé WSN-DS et il est destiné à aider les chercheurs à travailler sur les attaques Dos dans les RCSFs. Almomani et all, Ont utilisé ce dataset pour former un « Artificial Neural Network (ANN) » qui permet de détecteret de classer les types d'attaques Dos. Le protocole de routage LEACH a été utilisé pour recueillir l'ensemble de données, il est l'un des protocoles de routage les plus courants et les plusutilisés dans les RCSFs.

Le dataset contient un total de 374661 enregistrements et 23 attributs, cependant dans les fichiers CSV, plusieurs attributs ne sont pas utilisés, y compris RSSI, Distance maximale à CH, Distance moyenne à CH, Énergie actuelle **[49]**. On trouve seulement 19 attributs décrits dans le **« tableau 3.1 »**. Quatre classes représentent quatre attaques Dos qui sont : Blackhole (10049), Grayhole (14596), Flooding (3312) et TDMA (6638), et normal avec les 340066 enregistrements restants**. [48]**

#	Nom d'attribut	Description de l'attribut
1	Id	ID unique pour distinguer le nœud du capteur
2	Time	Temps de simulation actuel du nœud.
3	Is_CH	Un indicateur permettant de distinguer si le nœud est CH avec la valeur 1 ou un nœud normal avec la valeur 0.
4	Who CH	L'ID de CH dans la ronde actuelle
5	Dist_TO_CH	la distance entre le nœud et son CH dans la ronde courante
6	ADV_S	le nombre de messages publicitaires diffusés (broadcast) par CH aux nœuds.
7	ADV_R	le nombre de messages publicitaires CH reçus des CHs
8	JOIN_S	le nombre de messages de demande de jointure envoyés par les nœuds au CH
9	JOIN_R	le nombre de messages de demande de jointure reçus par le CH des nœuds.
10	SCH_S	le nombre de messages de diffusion de l'horaire de TDMA publicitaires envoyés aux nœuds.
11	SCH_R	le nombre de messages d'horaire de TDMA reçus des CH.
12	Rank	l'ordre de ce nœud dans le calendrier TDMA.
13	DATA_S	le nombre de paquets de données envoyés d'un capteur à son CH.
14	DATA_R	le nombre de paquets de données reçus de CH.
15	DATA_Sent_TO_BS	le nombre de paquets de données envoyés au BS
16	Dist_CH_TO_BS	la distance entre la CH et la BS.
17	Send_code	le code d'envoi du cluster.
18	Expaned energy	la quantité d'énergie consommée au cours de la ronde précédente.
19	Attack type	type du nœud. C'est une classe de cinq valeurs possibles, Blackhole, Grayhole, Flood, et TDMA, et normal, si le nœud n'est pas un attaquant

Table III-1: Description des attributs [39]

III.7 Protocol LEACH :

LEACH est un protocole de routage hiérarchique utilisé dans les WSN pour augmenter la durée de vie du réseau. Il suppose que la station de base est fixe et situé loin des nœuds de capteurs. Les nœuds sont homogènes et ont une énergie et une mémoire limitée. Les capteurs peuvent communiquer entre eux et ils peuvent communiquer directement avec la BS.

Le protocole LEACH garantit l'organisation des nœuds en clusters pour distribuer l'énergie entre eux, et dans chaque cluster il y a un nœud appelé Cluster Head (CH) qui collecte les données reçues de son cluster des membres et les transmettre à la BS.

Chaque cycle du protocole LEACH se compose principalement de deux phases :

- la phase de configuration : les clusters sont formés.
- La phase d'état stable: les données collectées seront transférées au nœud à la BS.

III.8 Expérience et évaluation:

Dans ce travail, on a appliqué cinq algorithmes de classification « SVM, DT, LR, KNN et Naïve Bayes » sur le dataset WSN-DS des attaques. Ceci afin de les évaluer sur la base des métriques d'évaluations définis dans la section précédente.

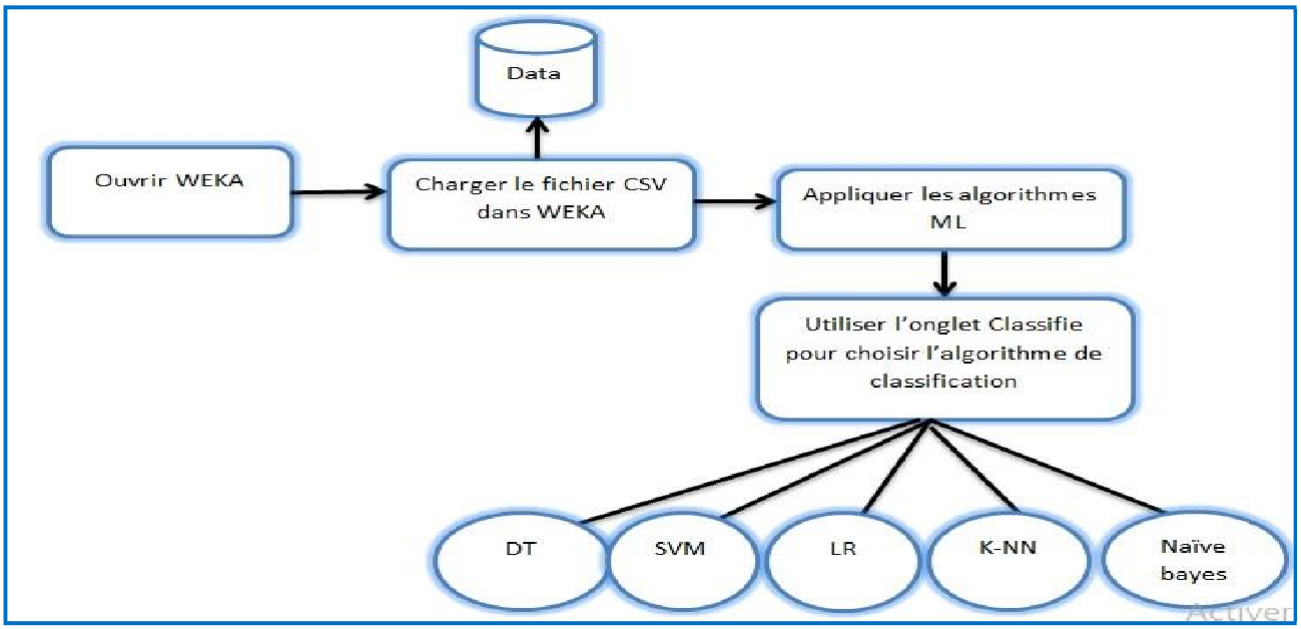

Figure III-2: Le workflow du modèle de classification

III.8.1 L'arbre de décision :

❖ Arbre de décision avec ensemble de données complet :

Cette section présente et examine les résultats de l'utilisation de l'arbre de décision J48 disponible dans WEKA. Les meilleurs résultats ont été obtenus avec **10-Fold Cross Validation**. Le résumé des résultats obtenus est présenté dans le tableau 3.2, l'exactitudedétaillée par classe est dans le tableau 3.3, et la matrice de confusion dans le tableau 3.4.ou le nombre de feuilles est de 316 et la taille de l'arbre est de 631. Le temps pris pour construire ce modèle est de

27.3 secondes.

Correctly Classified Instances	373401	99.6637 %
Incorrectly Classified Instances	1260	0.3363 %
Kappa statistic	0.9805	
Mean absolute error	0.002	
Root mean squared error	0.0351	
Relative absolute error	2.9202 %	
Root relative squared error	18.8321 %	
Total Number of Instances	374661	

Table III-2: Résumé des résultats de l'arbre décision obtenus

- **Correctly Classified Instances :** Le nombre d'exemples bien classés, en valeur absolue, puis en pourcentage du nombre total d'exemples**. [50]**

- **Incorrectly Classified Instances** : Sous le même format, le nombre d'exemples mal classés. **[50]**

- **Kappa statistic :** La statistique de Kappa (ou valeur) est une mesure qui compare l'exactitude observée à l'exactitude attendue (chance aléatoire).Il est utilisée non seulement pour évaluer un seul classificateur, mais aussi pour évaluer les classificateurs entre eux. **[50]**

- **Mean absolute error :** Erreur absolue en moyenne quantité utilisée pour mesurer la proximité des prévisions ou des prédictions par rapport aux résultats éventuels. **[50]**

- **Root mean squared error :** est l'écart-type des erreurs qui se produisent lorsqu'une prédiction est faite sur un dataset. **[50]**

- **Relative absolute error :** L'erreur absolue relative est un moyen de mesurer le rendement d'un modèle prédictif. Il est principalement utilisé dans l'apprentissage automatique, l'exploration de données. **[50]**

- **Root relative squared error :** L'erreur quadratique relative de la racine est relative à ce qu'elle aurait été si un simple prédicteur avait été utilisé. Plus précisément, ce simple prédicteur n'est que la moyenne des valeurs réelles**. [50]**

TP Rate	FP Rate	Precision	Rescall	F-Measure	MCC	ROC Area	PRC Area	Class
0,999	0,020	0,998	0,999	0,998	0,984	0,990	0,998	Normal
0,975	0,000	0,953	0,975	0,964	0,964	0,995	0,972	Flooding
0,927	0,000	0,995	0,927	0,960	0,960	0,966	0,935	TDMA
0,982	0,001	0,984	0,982	0,983	0,982	0,997	0,977	Grayhole
0,992	0,000	0,985	0,992	0,988	0,988	0,998	0,993	Blackhole
0,997	0,018	0,997	0,997	0,997	0,983	0,991	0,995	Weighted Avg

Table III-3: l'exactitude détaillée de l'arbre décision par classe

- **TP Rate :** représente le taux de cas d'attaque correctement identifié**. [48]**

  TPR=(TP)/ (TP+FN)

- **FP Rate :** représente le taux de cas sans attaques identifiés comme attaques par le système. **[48]**

  FPR=(FP)/ (FP+TN)

- **F-Measure**= (TP)/(TP+FP+FN) **[49]**

Normal	Flooding	TDMA	Grayhole	Blackhole	
a	b	c	d	e	<-- Classified as
339719	159	27	150	11	a= Normal
83	3229	0	0	0	b= Flooding
478	0	6151	5	4	c=TDMA
120	0	3	14332	141	d= Grayhole
2	0	1	76	9970	e=Blackhole

Table III-4: la matrice de confusion d'un l'arbre décision

❖ Arbre de décision avec un ensemble de données des attaques Sélectionnées :

Différents types d'attaques DOS peuvent affecter WSN à différents degrés. Zargar et autre **[51]** ont souligné que les attaque flooding et grayhole constituent les plus grandes menace pour la sécurité des RCSFs, car ils s'agissent des attaques les plus dangereuses, cela est dû au fait qu'ils sont très difficile à détecter.

Dans cette expérience, seules les attaques flooding et de grayhole sont incluses pour tester la précision du système .Le résume des résultats est présenté dans le tableau 3.5, l'exactitude détaillée par classe dans le tableau 3.6, et la matrice de confusion dans le tableau 3.7.Le nombre de feuilles est de 176 et la taille de l'arbre est de 351.Le temps nécessaire pour construire ce modèle est de 5.56 secondes.

Correctly Classified Instances	357446	99.8525 %
Incorrectly Classified Instances	528	0.1475 %
Kappa statistic	0.9846	
Mean absolute error	0.0012	
Root mean squared error	0.0298	
Relative absolute error	1.9175 %	
Root relative squared error	16.6605 %	
Total Number of Instances	357974	

Table III-5: Résumer des résultats de « DT » utilisant des attaques sélectionnées

TP Rate	FP Rate	Precision	Rescall	F-Measure	MCC	ROC Area	PRC Area	Class
0,999	0,012	0,999	0,999	0,999	0,985	0,997	1.000	Normal
0,974	0,000	0,955	0,974	0,964	0,964	0,995	0,968	Flooding
0,991	0,000	0,990	0,991	0,990	0,990	0,998	0,984	Grayhole
0,999	0,012	0,999	0,999	0,999	0,985	0,997	0,999	Weighted Avg

Table III-6: l'exactitude détaillée de « DT » par classe utilisant des attaques sélectionnées

a	b	c	<-- Classified as
339760	153	153	a= Normal
86	3226	0	b= Flooding
136	0	14460	c= Grayhole

Table III-7: la matrice de confusion d'un « DT » utilisant des attaques sélectionnées

❖ Comparaison entre l'ensemble de donnée complet avec des attaques « flooding, grayhole » :

Cette section compare l'ensemble de donnée complet et l'ensemble de donnée avec seulement les attaques flooding et grayhole en termes de nombre de feuilles et la taille de l'arbre « Figure3.3 », le temps nécessaire pour construire le modèle « Figure 3.4 », les instances correctement et incorrectement classées sont illustres dans « Figure 3.5 ».

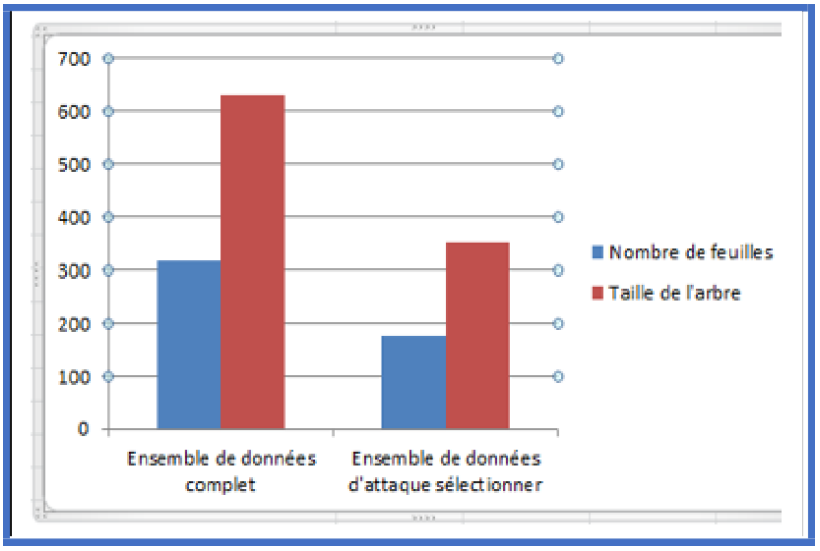

Figure III-3 : DT Comparaison en terme de nombre de feuille et la taille de l'arbre

Analyse 1: Dans la figure III-3 le nombre de feuilles et la taille de l'arbre dans le cas d'un dataset complet avec (316 feuilles, 631) est supérieur que dataset sélectionner avec (176 feuilles, 351).

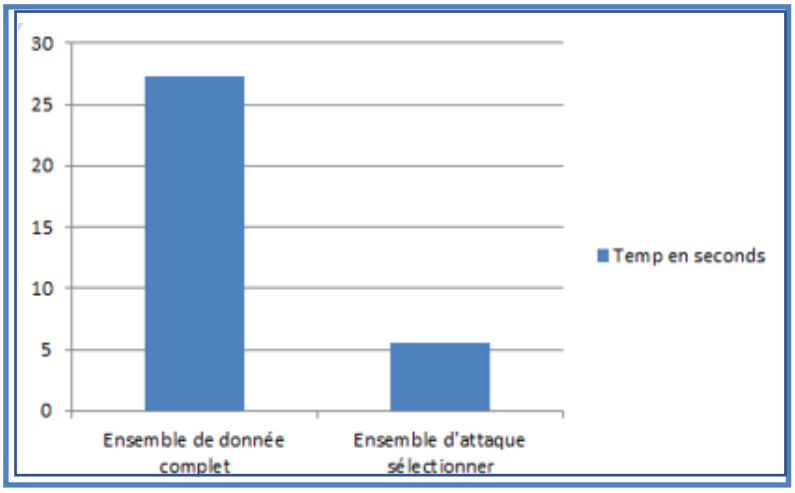

Figure III-4: DT Comparaison en terme de temps

Analyse 2: Dans le cas d'un dataset sélectionner, le classificateur DT est plus rapide à construire le modèle avec seulement 5.56 secondes, mais dans le cas d'un dataset complet a mis le plus de temps pour construire à 27.3 secondes, puisqu'il contient un grand nombre d'enregistrement.

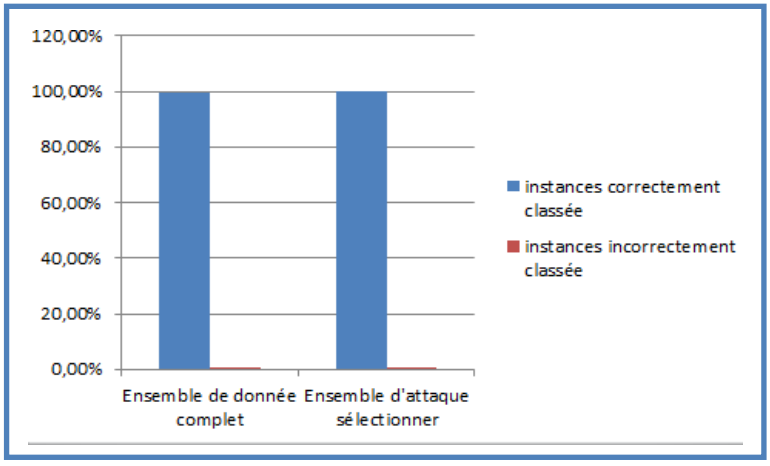

Figure III-5 : DT Comparaison en terme des instances correctement et incorrectement Classées

Analyse 3: Dans la figure III-5 les instances correctement et incorrectement classés dans le cas d'un dataset sélectionné et dataset complet sont proches.

Conclusion : à partir de ces résultats on peut remarquer que dans l'algorithme de classification DT,l'utilisation de l'ensemble de données d'attaque sélectionné est meilleure que l'utilisation de l'ensemble de données complet en termes de nombre de feuilles et la taille de l'arbre, temps nécessaire pour construire le modèle et le nombre d'Instances correctement et incorrectement classées.

III.8.2 Machine vectorielle de soutien:

❖ Machine vectoriel de soutien avec ensemble de données complet :

Cette section présente et examine les résultats de l'utilisation de la SVM avec un mode de test de validation croisée à dix niveaux. Le résume des résultats obtenus est présenté dans le tableau 3.8, l'exactitude détaillée par classe est dans le tableau 3.9 et la matrice de confusion dans le tableau 3.10. Le temps nécessaire pour construire ce modèle est de 612.39 secondes.

Correctly Classified Instances	363831	97.1094%
Incorrectly Classified Instances	10830	2.8906%
Kappa statistic	0.8313	
Mean absolute error	0.2412	
Root mean squared error	0.318	
Relative absolute error	347.504%	
Root relative squared error	170.708%	
Total Number of Instances	374661	

Table III-8 : Résumer des résultats de la SVM obtenus

TP Rate	FP Rate	Precision	Rescall	F-Measure	MCC	ROC Area	PRC Area	Class
0,994	0,087	0,991	0,994	0.993	0.920	0.956	0.991	Normal
0,941	0,001	0,902	0,941	0.921	0.920	0.999	0.890	Flooding
0,862	0,001	0,941	0,862	0.900	0.899	0.932	0.814	TDMA
0,501	0,005	0,802	0,501	0.617	0.623	0.980	0.592	Grayhole
0,955	0,015	0,643	0,955	0.769	0.777	0.992	0.631	Blackhole
0,971	0,079	0,973	0,971	0.970	0.904	0.958	0.962	Weighted Avg

Table III-9: l'exactitude détaillée de la SVM par classe

a	b	c	d	e	<-- Classified as
338084	339	146	1478	10	a= Normal
172	3116	1	0	23	b= Flooding
870	0	5724	3	41	c=TDMA
1957	0	75	7307	5257	d= Grayhole
0	0	138	311	9600	e=Blackhole

Table III-10 : la matrice de confusion d'un SVM

❖ Machine vectoriel de soutien avec un ensemble de données des attaques sélectionnées :

Dans cette expérience, seules les attaques flooding et de grayhole sont incluses pour tester la précision du système .Le résume de résultats est présenté dans le tableau 3.11, l'exactitude détaillée par classe est dans le tableau 3.12, et la matrice de confusion dans le tableau 3.13.Le nombre de feuilles est de 176 et la taille de l'arbre est de 351.Le temps nécessaire pour construire ce modèle est de 80.46 secondes.

Correctly Classified Instances	356522	99.5944 %
Incorrectly Classified Instances	1452	0.4056 %
Kappa statistic	0.9581	
Mean absolute error	0.2232	
Root mean squared error	0.2739	
Relative absolute error	349.3763 %	
Root relative squared error	153.2519 %	
Total Number of Instances	357974	

Table III-11: résumer des résultats de « SVM » utilisant des attaques sélectionnées

TP Rate	FP Rate	Precision	Rescall	F-Measure	MCC	ROC Area	PRC Area	Class
0,997	0,028	0,999	0,997	0,998	0,959	0,985	0.998	Normal
0,939	0,001	0,903	0,939	0,921	0,920	0,995	0,857	Flooding
0,978	0,002	0,958	0,978	0,968	0,967	0,989	0,938	Grayhole
0,996	0,027	0,996	0,996	0,996	0,959	0,985	0,995	Weighted Avg

Table III-12: l'exactitude détaillée de « SVM » par classe utilisant des attaques Sélectionnées

a	b	c	<-- Classified as
339140	335	591	a= Normal
174	3111	27	b= Flooding
325	0	14271	c= Grayhole

Table III-13: la matrice de confusion d'un « SVM » utilisant des attaques sélectionnées

❖ Comparaison entre l'ensemble de donnée complet et l'ensemble des attaques sélectionnées :

Cette section compare Dataset complet et Dataset avec seulement les attaques flooding et grayhole en termes le temps nécessaire pour construire le modèle « Figure 3.6 », les instances correctement et incorrectement classées « Figure 3.7 ».

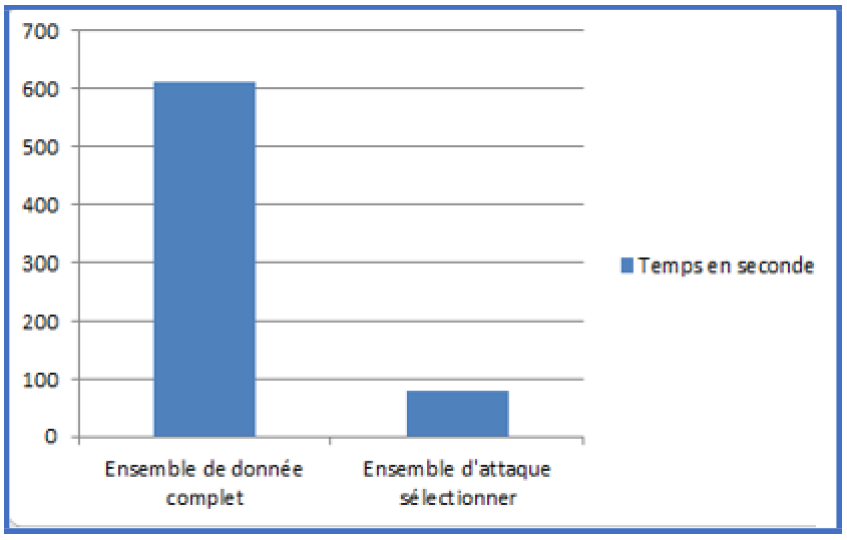

Figure III-6: SVM Comparaison en terme de temps

Analyse 1: dans le cas d'un dataset sélectionner, le classificateur SVM est plus rapide à construire le modèle avec 80.46 secondes, mais dans le cas d'un dataset complet a mis le plus beaucoup de temps pour construire à 612.39 secondes.

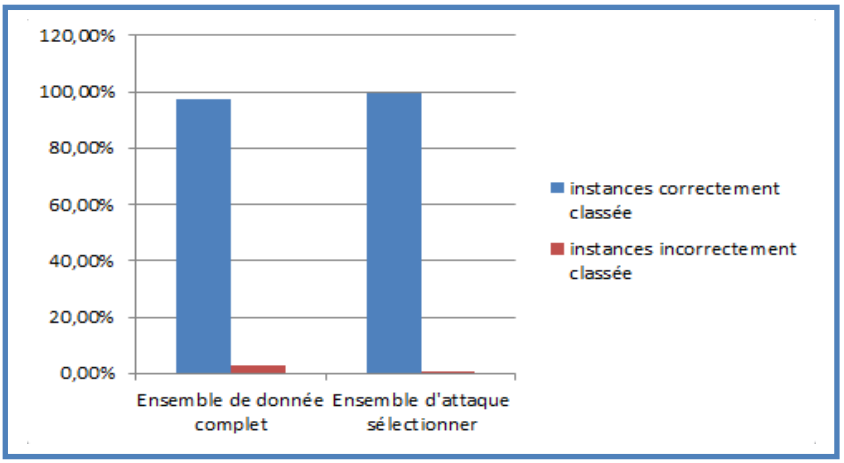

Figure III-7: SVM Comparaison en terme des instances correctement et incorrectement classées

Analyse 2: Le classificateur SVM avait les instances correctement classée la plus élevée dans le cas d'un dataset sélecctionner (99.59%), ce qui correspond à une diminution des instances incorrectement classés (0.40%). Tandis que le classificateur dans le cas d'un dataset complet avait la plus faible en termes des instances correctement classés (97.10%), ce qui correspond à une hausse des instances incorrectement classés (2.89%).

Conclusion : à partir de ces résultats on peut remarquer que dans l'algorithme de classification SVM,l'utilisation de l'ensemble de données des attaques sélectionnées est meilleure que l'utilisation del'ensemble de données complet en termes de temps nécessaire pour construire le modèle et le nombre d'Instances correctement et incorrectement classées.

III.8.3 Régression logistique :

❖ Régression logistique avec ensemble de données complet :

Cette section présente et examine les résultats de l'utilisation de régression logistique disponible dans WEKA. Le résume des résultats obtenus est présenté dans le tableau 3.14, l'exactitude détaillée par classe est dans le tableau 3.15, et la matrice de confusion dans le tableau 3.16. Avec le temps de construction du modèle : 303.88 secondes.

Correctly Classified Instances	365504	97.5559 %
Incorrectly Classified Instances	9157	2.4441 %
Kappa statistic	0.8595	
Mean absolute error	0.0122	
Root mean squared error	0.0768	
Relative absolute error	17.647 %	
Root relative squared error	41.2284 %	
Total Number of Instances	374661	

Table III-14: Résumer des résultats de la LR obtenus

TP Rate	FP Rate	Precision	Rescall	F-Measure	MCC	ROC Area	PRC Area	Class
0,995	0,042	0,996	0,995	0,995	0,951	0,993	0,999	Normal
0,915	0,001	0,903	0,909	0,909	0,908	1,000	0,950	Flooding
0,918	0,000	0,980	0,948	0,948	0,948	0,971	0,937	TDMA
0,688	0,009	0,749	0,717	0,717	0,707	0,994	0,809	Grayhole
0,785	0,011	0,669	0,722	0,722	0,716	0,995	0,847	Blackhole
0,976	0,039	0,976	0,976	0,976	0,935	0,993	0,986	Weighted Avg

Table III-15: l'exactitude détaillée de la LR par classe

a	b	c	d	e	<-- Classified as
338447	324	99	1190	6	a= Normal
281	3031	0	0	0	b= Flooding
517	0	6097	17	7	c=TDMA
635	0	26	10043	3892	d= Grayhole
10	0	0	2153	7886	e=Blackhole

Table III-16• : la matrice de confusion d'un LR

❖ Régression logistique avec un ensemble de données des attaques sélectionnées :

Dans cette expérience, seules les attaques flooding et de grayhole sont incluses pour tester la l'exactitude du système .Le résumer de résultats est présenté dans le tableau 3.17, l'exactitude détaillée par classe dans le tableau 3.18, et la matrice de confusion dans le tableau 3.19.Le nombre de feuilles est de 176 et la taille de l'arbre est de 351.Le temps nécessaire pour construire ce modèle est de 42.99secondes.

Correctly Classified Instances	355695	99.3634%
Incorrectly Classified Instances	2279	0.6366%
Kappa statistic	0.9344	
Mean absolute error	0.0076	
Root mean squared error	0.0583	
Relative absolute error	11.8764 %	
Root relative squared error	32.6079 %	
Total Number of Instances	357974	

Table III-17: Résumer des résultats de « LR » utilisant des attaques sélectionnées

TP Rate	FP Rate	Precision	Rescall	F-Measure	MCC	ROC Area	PRC Area	Class
0,9976	0,050	0,997	0,996	0,997	0,934	0,998	1,000	Normal
0,916	0,001	0,905	0,916	0,910	0,909	1,000	0.950	Flooding
0,958	0,003	0,929	0,958	0,943	0,941	0,999	0,927	Grayhole
0,994	0,047	0,994	0,994	0,994	0,934	0,998	0,996	Weighted Avg

Table III-18: l'exactitude détaillée de « LR » par classe utilisant des attaques sélectionnées

a	b	c	<-- Classified as
338674	320	1072	a= Normal
279	3033	0	b= Flooding
608	0	14271	c= Grayhole

Table III-19: la matrice de confusion d'un « LR » utilisant des attaques sélectionnées

- **Comparaison entre l'ensemble de donnée complet et l'ensemble des attaques sélectionnées :**

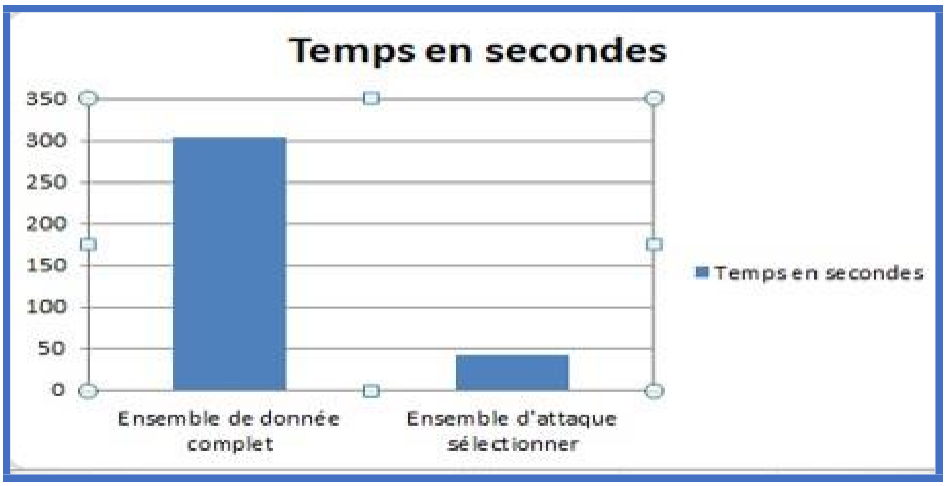

Figure III-8: LR Comparaison en terme de temps

Analyse 1: dans le cas d'un dataset sélectionner, le classificateur LR est plus rapide à construire le modèle avec seulement 42.99 secondes, mais dans le cas d'un dataset complet a mis le plus beaucoup de temps pour construire à 303.88 secondes.

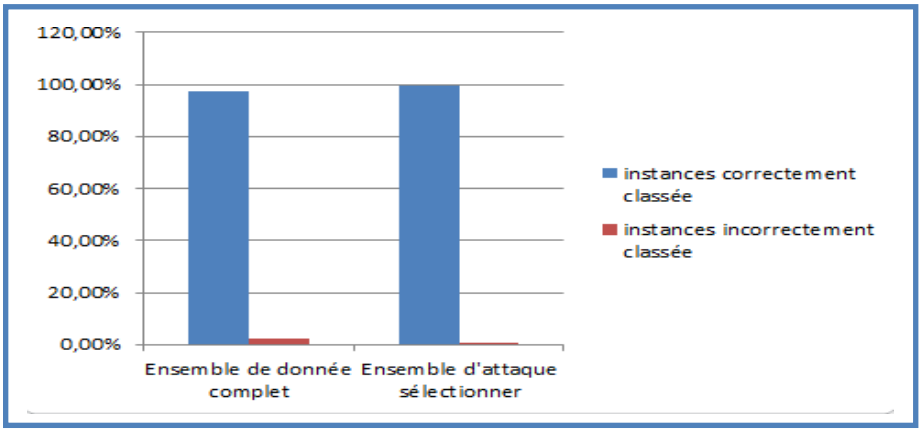

Figure III-9: LR Comparaison en termes d'instances correctement et incorrectement Classées

Analyse 2: Le classificateur LR avait les instances correctement classée la plus élevée dans le cas d'un dataset sélecctionner (99.36%), ce qui correspond à une diminution des instances incorrectement

classés (0.63%). Tandis que le classificateur dans le cas d'un dataset complet avait la plus faible en termes des instances correctement classés (97.55%), ce qui correspond à une hausse des instances incorrectement classés (2.44%).

Conclusion : d'après ces résultats nous pouvons voir que dans la classification LR, l'utilisation de l'ensemble de données d'attaque sélectionné est meilleure que l'utilisation de l'ensemble de données complet en termes de temps nécessaire pour construire le modèle et le nombre d'Instances correctement et incorrectement classées.

III.8.4 K-Plus Proches Voisins :

❖ K-plus proches voisins avec ensemble de données complet :

Cette section présente et examine les résultats de l'utilisation de K-plus proches voisins disponible dans WEKA. Le résumer des résultats obtenus est présenté dans le tableau 3.20, l'exactitude détaillée par classe dans le tableau 3.21, et la matrice de confusion dans le tableau 3.22.Le temps nécessaire pour construire ce modèle est de 0.22 secondes.

Correctly Classified Instances	372593	99.448%
Incorrectly Classified Instances	2068	0.552%
Kappa statistic	0.9682	
Mean absolute error	0.0022	
Root mean squared error	0.047	
Relative absolute error	3.1883 %	
Root relative squared error	25.2214 %	
Total Number of Instances	374661	

Table III-20 : Résumer des résultats de la K-NN obtenus

TP Rate	FP Rate	Precision	Rescall	F-Measure	MCC	ROC Area	PRC Area	Class
0.997	0,028	0,997	0,997	0,997	0,970	0,985	0,997	Normal
0.922	0,001	0,931	0,922	0,927	0,926	0.960	0,861	Flooding
0.920	0,001	0,926	0,920	0,923	0,922	0.960	0,855	TDMA
0.981	0,001	0,979	0,981	0,980	0,979	0.990	0,960	Grayhole
0.992	0,000	0,991	0,992	0,922	0,992	0,996	0,984	Blackhole
0.994	0,025	0,994	0,994	0,994	0,970	0,985	0,992	Weighted Avg

Table III-21: l'exactitude détaillée de la K-NN par classe

a	b	c	d	e	<-- Classified as
339137	225	468	213	5	a= Normal
249	3055	0	7	1	b= Flooding
513	0	6109	11	5	c=TDMA
190	2	7	14321	76	d= Grayhole
3	0	11	64	9971	e=Blackhole

Table III-22: la matrice de confusion d'un K-NN

❖ K-plus proches voisins avec un ensemble de Données des attaques sélectionnées :

Dans cette expérience, seules les attaques flooding et de grayhole sont incluses pour tester la précision du système. Le résumer de résultats est présenté dans le tableau 3.23, l'exactitude détaillée par classe dans le tableau 3.24, et la matrice de confusion dans le tableau 3.25.Le nombre de feuilles est de 176 et la taille de l'arbre est de 351.Le temps nécessaire pour construire ce modèle est 0.19 secondes.

Correctly Classified Instances	357061	99.745%
Incorrectly Classified Instances	913	0.255%
Kappa statistic	0.9734	
Mean absolute error	0.0017	
Root mean squared error	0.0412	
Relative absolute error	2.6685 %	
Root relative squared error	23.0748 %	
Total Number of Instances	357974	

Table III-23: Résumer des résultats « KNN » utilisant des attaques sélectionnées

TP Rate	FP Rate	Precision	Rescall	F-Measure	MCC	ROC Area	PRC Area	Class
0,999	0,025	0,999	0,999	0,999	0,973	0,987	0.999	Normal
0,921	0,001	0,932	0,921	0,921	0,926	0,964	0,858	Flooding
0,987	0,001	0,989	0,987	0,987	0,984	0,993	0,968	Grayhole
0,997	0,024	0,997	0,997	0,997	0,973	0,987	0,996	Weighted Avg

Table III-24: l'exactitude détaillée de « KNN » par classe utilisant des attaques sélectionnées

a	b	c	<-- Classified as
339608	219	239	a= Normal
225	3051	6	b= Flooding
192	2	14402	c= Grayhole

Table III-25: la matrice de confusion d'un « KNN » utilisant des attaques sélectionnées

- **Comparaison entre l'ensemble de donnée complet et l'ensemble des attaques sélectionnées :**

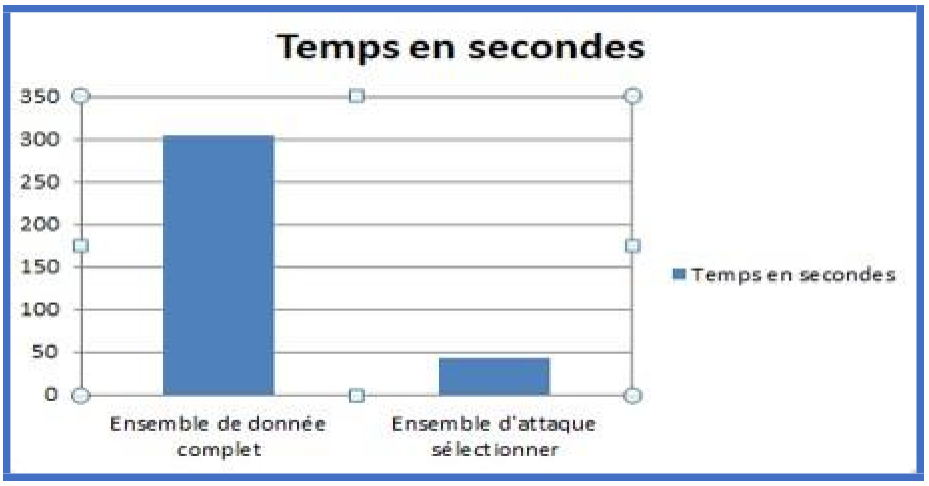

Figure III-10: K-NN Comparaison en terme de temps

Analyse 1: dans le cas d'un dataset sélectionner, le classificateur LR est plus rapide à construire le modèle avec seulement 0.19 secondes, mais dans le cas d'un dataset complet a mis le plus de temps pour construire à 0.22 secondes.

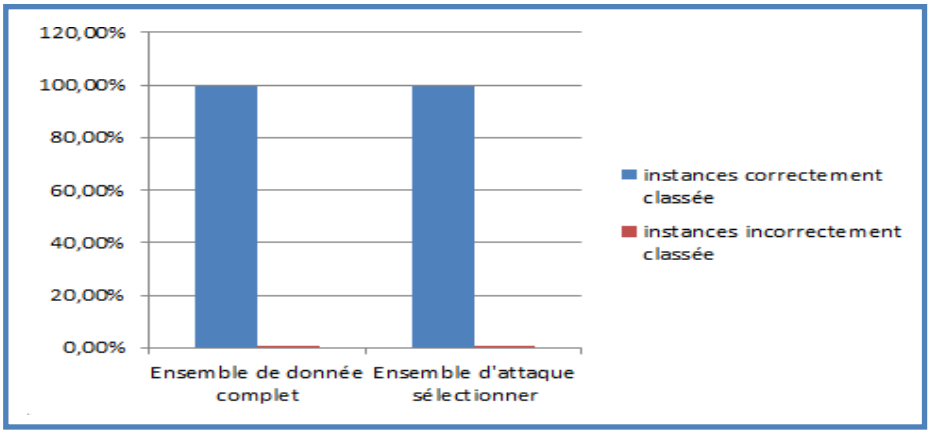

Figure III-11: K-NN Comparaison en terme d'instances correctement et incorrectement Classées

Analyse 2: Dans la figure III-11 les instances correctement et incorrectement classés dans le cas d'un dataset sélectionné et dataset complet sont proches.

Conclusion : en fonction de ces résultats on peut noter que dans la classification K-NN, l'utilisation de l'ensemble de données d'attaque sélectionné est meilleure que l'utilisation de l'ensemble de données complet en termes de temps nécessaire pour construire le modèle et le nombre d'Instances correctementet incorrectement classées.

III.8.5 Bayes Naïve :

❖ Bayes Naïve avec ensemble de données complet :

Cette section présente et examine les résultats de l'utilisation de Naïve Bayes disponible dans WEKA. Le résumer des résultats obtenus est présenté dans le tableau 3.26, l'exactitude détaillée par classe dans le tableau 3.27, et la matrice de confusion dans le tableau 3.28. Avec le temps de construction du modèle : 1.82 secondes.

Correctly Classified Instances	357327	95.3734 %
Incorrectly Classified Instances	17334	4.6266 %
Kappa statistic	0.7631	
Mean absolute error	0.0186	
Root mean squared error	0.1334	
Relative absolute error	26.7746 %	
Root relative squared error	71.623 %	
Total Number of Instances	374661	

Table III-26: Résumer des résultats de la Naïve Bayes obtenus

TP Rate	FP Rate	Precision	Rescall	F-Measure	MCC	ROC Area	PRC Area	Class
0,972	0,012	0,999	0,972	0,985	0,864	0,980	0,997	Normal
1,000	0,010	0,469	1,000	0,638	0,681	1,000	0,903	Flooding
0,755	0,000	0,988	0,755	0,856	0,862	0,956	0,834	TDMA
0,589	0,018	0,571	0,589	0,580	0,563	0,983	0,621	Grayhole
0,993	0,018	0,601	0,993	0,749	0,765	0,992	0,697	Blackhole
0,954	0,012	0,967	0,954	0,958	0,848	0,980	0,971	Weighted Avg

Table III-27: l'exactitude détaillée de la Naïve Bayes par classe

a	b	c	d	e	<-- Classified as
330421	3200	45	6361	39	a= Normal
0	3311	0	1	0	b= Flooding
423	242	5009	41	923	c=TDMA
0	311	12	8604	5669	d= Grayhole
0	0	2	65	9982	e=Blackhole

Table III-28: la matrice de confusion d'un Naïve Bayes

❖ Bayes Naïve avec un ensemble de données des attaques sélectionnées :

Dans cette expérience, seules les attaques flooding et de grayhole sont incluses pour tester la précision du système .Le résumer de résultats est présenté dans le tableau 3.29, l'exactitude détaillée par classe dans le tableau 3.30, et la matrice de confusion dans le tableau 3.31.Le nombre de feuilles est de 176 et la taille de l'arbre est de 351.Le temps nécessaire pour construire ce modèle est de 1.34 Secondes.

Correctly Classified Instances	347987	97.2101%
Incorrectly Classified Instances	9987	2.7899%
Kappa statistic	0.7689	
Mean absolute error	0.0188	
Root mean squared error	0.133	
Relative absolute error	29.3998 %	
Root relative squared error	74.4158 %	
Total Number of Instances	357974	

Table III-29: Résumer des résultats de «naïve bayes » utilisant des attaques sélectionnées

TP Rate	FP Rate	Precision	Rescall	F-Measure	MCC	ROC Area	PRC Area	Class
0,972	0,000	1.000	0,972	0,986	0,794	0,984	0.999	Normal
1.000	0,010	0.486	1.000	0,654	0,694	0,999	0,846	Flooding
0,979	0,019	0,688	0,979	0,808	0,812	0,995	0,812	Grayhole
0,972	0,001	0,983	0,972	0,975	0,794	0,984	0,990	Weighted Avg

Table III-30: l'exactitude détaillée de « naïve bayes » par classe utilisant des attaques Sélectionnées

a	b	c	<-- Classified as
330386	3197	6483	a= Normal
0	3311	1	b= Flooding
0	306	14290	c= Grayhole

Table III-31: la matrice de confusion d'un « naïve bayes » utilisant des attaques sélectionnées

❖ **Comparaison entre l'ensemble de donnée complet et l'ensemble des attaques sélectionnées :**

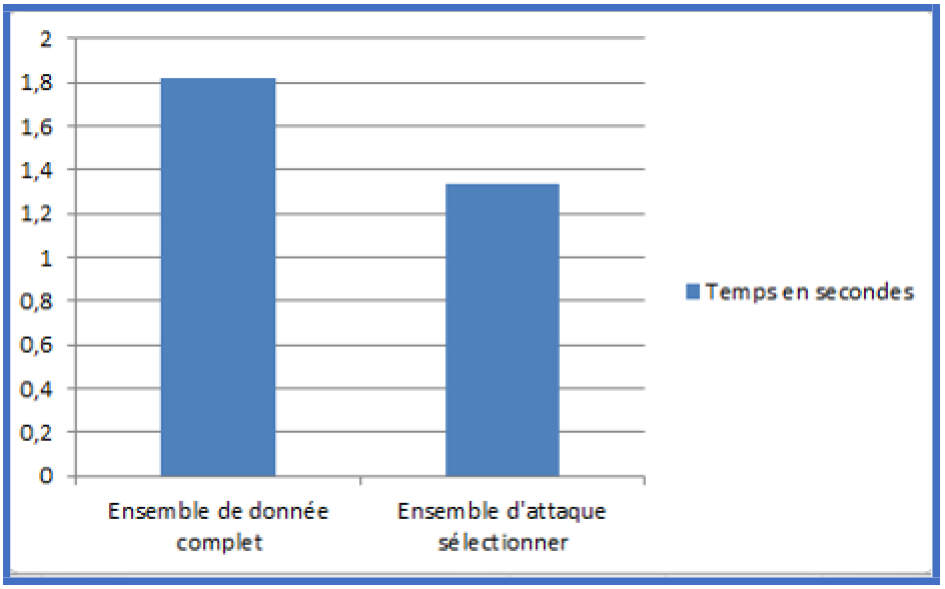

Figure III-12: naïve bayes Comparaison en terme de temps

Analyse : dans le cas d'un dataset sélectionner, le classificateur naïve bayes est plus rapide à construire le modèle avec seulement 1.34 secondes, mais dans le cas d'un dataset complet a mis le plus de temps pour construire à 1.82 secondes.

III.8 Résultat :

Dans le but de mesures les performances des classificateur choisis. La figure « 3.13» illustre lesrésultats comparatifs entre les classificateurs **« DT, SVM, K-NN, LR, naïve bayes »** en termes des instances correctement classées est dans le cas d'un ensemble de donnée des attaque sélectionnés.

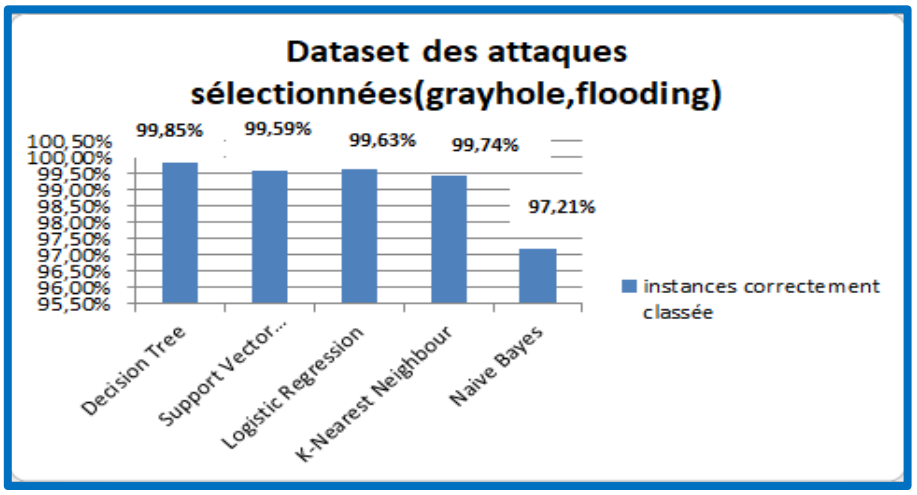

Figure III-13: comparaison entre les algorithmes en terme des instances correctement Classées

Analyse : classificateur Naive Bayes avait des instances correctement classée la plus faible à 97.2%, tandis que l'arbre de décision (Decision Tree) avait des instances correctement classée la plus élevée à 99.85%, K-Nearest Neighbour arrivé en deuxième position avec 99.74%, et Logistic Regression a atteint 99.63%.Support Vector Machine avait des instances correctement classée de 99.59%.

Conclusion Générale

Conclusion générale :

Dans ce travail, nous avons évalué les techniques de classification de l'apprentissage automatique pour la détection des attaques DOS dans les RCSFs. Un dataset spécialisées pour les RCSFs appelé WSN-DS a été utilisé pour comparer les performances des techniques d'apprentissage automatique.

Au début, nous avons eu un peu de difficulté à trouver ce dataset parce que les chercheurs qui les ont créés n'ont pas répondu à nos messages. Mais après une longue recherche, nous avons trouvé ce dataset dans un site « Kaggle ».ce dataset a été utilisé pour classer quatre types d'attaques DOS : blackhole, grayhole, flooding et TDMA. Différentes techniques de ML ont été testées:« DT, SVM, LR, K-NN, Naïve Bayes ». WEKA a utilisé pour tester les performances de la détection DoS sur WSN-DS, les meilleurs résultats ont été obtenus avec 10-Fold Cross Validation. L'ensemble de données complet a été utilisé et un ensemble de données réduit avec seulement les attaques flooding et de grayhole les plus dangereuses a alors été considéré. Nous avons constaté que l'arbre de décision(DT) obtenait de meilleurs résultats de classification que d'autres techniques dans le cas d'un ensemble de données d'attaque sélectionné, avec Correctly Classified Instances égal 99.8525 %.

À l'avenir, cette recherche sera étendue à d'autres types de classificateurs et de techniques d'apprentissage automatique. Il est également possible pour les recherches futures d'envisager d'autres scénarios d'attaque, ou des attaques sur des protocoles autres que LEACH.

Références Bibliographiques

Références bibliographiques :

[1]N. Marriwala and P. Rathee, “An approach to increase the wireless sensor network lifetime,” in Proceedings of the World Congress on Information andCommunication Technologies (WICT ’12), pp. 495–499, IEEE, Trivandrum,India, OctoberNovember 2012.

[2]V. C. Gungor, B. Lu, and G. P. Hancke, “Opportunities and challenges of wireless sensor networks in smart grid,” IEEE Transactions on IndustrialElectronics, vol. 57, no. 10, pp. 3557– 3564, 2010

[3]I. Butun, S. D. Morgera, and R. Sankar, “A survey of intrusion detection systems in wireless sensor networks,” IEEE Communications Surveys &Tutorials, vol. 16, no. 1, pp. 266–282, 2014.

[4]Boubiche Djallel Eddin. Une approche Inter-Couches (cross-layer) pour laSécurité dans les R.C.S.F. Thèse de doctorat. Batna.

[5]M. A. Rassam, M. A. Maarof, and A. Zainal, “A survey of intrusion detectionschemes in wireless sensor networks,” American Journal of AppliedSciences,vol. 9, no. 10, pp. 1636–1652, 2012.

[6] Lama Alsulaiman, Saad Al-Ahmadi. «PERFORMANCE EVALUATION OFMACHINE LEARNING TECHNIQUES FOR DOS DETECTION IN WIRELESS SENSOR NETWORK." Internationel Journal of NetworkSecurity & Its Application, Vol 13, No 2, March 2021.

[7] « Capteur - Définition et Explications », sur le site « techno-science.net », consulter le 30/05/2022, Capteur : définition et explications (techno-science.net).

[8] «Les capteurs », sur le site Science de l’ingénieur, consulter le 30/05/2022, Les capteurs – Sciences de l'Ingénieur (blaisepascal.fr).

[9] DAKIRA FORMATION. *Les capteurs* شرح رائع ومميز حول أجهزة الاستشعار [Vidéo],consulter le 01/03/2022 https://youtu.be/M4VeF04_2D0.

[10] « Réseau de capteurs sans fil », sur le site Wikipédia, consulter le 21/04/2022, Réseau de capteurs sans fil — Wikipédia (wikipedia.org)

[11] Mohamed, FEHAM. Mise en place d'un réseau de capteurs sans fil pour la détection des feux de forét.Université de Tlemcen, Octobre 2013.

[12] HAFIR, Latifa et Radhia SLIMANI. Etude et évaluation des perfermances des protocoles de routage pour les réseaux de capteur sans fil. Thése de Master. Béjaia, 2015/2016.

[13] Deep learning خطرناكنو," chapitre 1-RCSF", sur le site SCRIBD. Consulter le 05/03/2022 Chapitre 1-RCSF | PDF | Réseau de capteurs sans fil |Réseau ad hoc sans fil (scribd.com).

[14] KABOU Salaheddine, Belgourari Abdessamed.Etat de l'art sur les réseaux de capteurs sans fil. Thèse de licence, Béchar,juin 2010.

[15] « Cancer : un capteur pour surveiller la tumeur et adapter le traitement en temps réel » sur le site Top Santé, consulter le 1/07/2022 Cancer : un capteur pour surveiller la tumeur et adapter le tr... - Top Santé (topsante.com).

[16] « Un capteur portable, il est facile de vérifier le niveau d'oxygéné et le pouls dans le corps humains », sur le site alamy, consulter le 11/05/2022 Un capteur portable, il est facile de vérifier le niveau d'oxygène et le pouls dans le corps humain Photo Stock - Alamy (alamyimages.fr).

[17] « Station météorologique sur batterie », sur le site Agri EXPO, consulter le 1/07/2022 Station météorologique sur batterie - Tous les fabricants de l'agriculture (agriexpo.online).

[18] BERRAZEG Khaled & BELAIDI Yamina. Détection des feux de forêt par réseau de capteurs. Thèse de master. Tlemcen, 2016/2017.

[19] « Domotique : tout comprendre aux protocoles pour la maison connectée »sur le site RANDROID, consulter le 1/06/2022 Domotique : tout comprendre aux protocoles pour la maison connectée (frandroid.com).

[20] « Stockage –Dépotage »sur le site FILTRANS, consulter le 1/06/2022 Stockage et Dépotage - Filtrans.

[21] BENBRAHIM, Salah-Eddine. DEFENSE CONTRE L'ATTAQUE D'ANALYSE DE TRAFIC DANS LES RESEAUX DE CAPTEUR SANSFIL(WSN). Thèse de master.Montréal, AOUT 2011.

[22] Idres Louiza, Système de détection d'intrusion hybride et hiérarchique pour les réseaux de capteur sans fil. Thèse de Master, Tizi-Ouzou, 2011,2012

[23] David Martins, Hervé Guyennet. État de l'art - Sécurité dans les réseauxde capteurs sans fil. SAR-SSI 2008: 3rd conference on Security of Network Architectures and Information Systems, 2008, France. Pp.167–181.

[24] Mémoire de fin d'études Pour l'obtention du diplôme de Master en Informatique/ Université Abou Bakr Belkaid– Tlemcen/ Option : Réseaux et Systèmes Distribués (R.S.D)Option : Réseaux et Systèmes Distribués (R.S.D)/ Utilisation del'apprentissage automatique pour la sécurité d'un réseau de radio cognitive/ réalisé par GHENNANI Hind Selma et MEDJDOUB Wissam /2017-2018

[25] Le contrôle d'accès, consulté le : 14/02/2022 https://www.locken.fr/base-de-connaissances-produits/qu-est-ce-que-le- controle-d-acces/

[26] Les principes de sécurité informatique , Consulter le : 03/04/2022 https://apcpedagogie.com/les- principes-de-securite-informatique/ http://pagesperso.univ-brest.fr/~bounceur/ecole2013/pdf_presentations/11_presentation_bensaber.pdf

[27] Mémoire de magistère dans Protocoles pour la Sécurité des Réseaux de Capteurs Sans Fil/ Université Hadj Lakhder-Batna/ réalise par Athmani samir /15-07-2018

[28] Différence entre attaque active et attaque passive, consulter le : 29/04/2022 https://waytolearnx.com/2018/07/difference-entre-attaque-active-et- attaque-passive.html

[29] Mémoire de Magistère en Informatique/Option : Réseaux et Systèmes Distribués/Thème:Sécurité dans les Réseaux de Capteurs Sans-Fil/Présenté par/Messai Mohamed Lamine/promo 2007/2008

[30] Consulter le 01/04/2022 Qu'est-ce qu'une collision de données? - définition de techopedia - Développement – 2022 consulter le 01/04/2022 https://fr.theastrologypage.com/data-collision

[31] Consulter le 01/04/2022 https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26557430

[32] Epuisement consulter 4/04/2022 https://www.cloudflare.com/fr-fr/learning/network-layer/what-is-the-

network-layer

[33] État de l'art - Sécurité dans les réseaux de capteurs sans fil David Martins, Hervé Guyennet/ HAL open sience / HAL Id: hal-00661898 https://hal.archives- ouvertes.fr/hal-00661898 Submitted on 20 Jan 2012.

[34] Mémoire de fin d'études/ Université Mouloud Mammeri de Tizi-Ouzou/Thème Système de détection d'intrusion hybrideet hiérarchique pour les réseaux de capteur sans fil /Réalisé par : Melle Idres Louiza/promo 2011-2012

[35] Mémoire présenté pour l'obtention Du diplôme de Master Académique /UNIVERSITE MOHAMED BOUDIAF - M'SILA/ sur le théme La Sécurité des Communications dans les Réseauxde Capteurs sans Fils/Par: Doumi Abdelmoumain/ Soutenu le: 24 /06 /2018

[36] Système de détection d'intrusion consulter le 30/04/2022 :,https://fr.wikipedia.org/wiki/Syst%C3%A8me_de_d%C3%A9tection_d%2 7intrusion

[37] Consulter le : 30/04/2022https://www.techno-science.net/glossaire-definition/Systeme-de-detection d-intrusion.html

[38] Système de détection d'intrusion Consulter le : 30/04/2022Q Système de détection d'intrusion - Définition et Explications (techno-science.net)

[39] BASSAM KASASBEH, « WSN-DS », sur le site Kaggle, consulter le07/01/2022, WSN-DS | Kaggle

[40] Kateb, N. (2011). Une approche multi agents pour le data mining. Thésemaster, Oum El Bouaghi.

[41] " K Nearest Neighbor (KNN) Algorithm for machine learning" ,sur lesite Java T Point,consulter le1/06/2022 K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint

[42] «Naïve Bayes Classifier Algorithm», sur le site Java T Point, consulterle1/06/2022 Naive Bayes Classifier in Machine Learning - Javatpoint.

[43] « Decision Tress Classification Algorithm», sur le site Java T Point, consulter le 1/06/2022, Machine Learning Decision Tree ClassificationAlgorithm - Javatpoint.

[44] Battah, M. (2020, janvier). Machine Learning Algorithms - A Review.International Journal of Science and Research (IJSR), 9, 6.

[45] «Logistic Regression in machine learning», sur le site Java T Point, consulter le 1/06/2022, Logistic Regression in Machine Learning - Javatpoint

[46] Al-isaa, A., Al-Akhras, M., Al sahli, M., & Alwairdhi, M. (2019). Using Machine Learning to Detect DoS Attacks in Wireless sensor network.IEEE jordan internationnal joint conference on electrical Engineering and informationTechnologie(JEEIT), (p. 6). Riyadh.

[47] Al-Akhras1 , M Alawairdhi1 , Al Alkoudari & S Atawne," USING MACHINE LEARNING TO BUILD A CLASSIFICATION MODEL FOR IOT NETWORKS TO DETECT ATTACK SIGNATURES", International Journal of Computer Networks & Communications (IJCNC) Vol.12, No.6, November 2020,Jordan

[48] I. Almomani, B. Al-Kasasbeh, and M. Al-Akhras, "WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks."Journal of Sensors, 2016. doi: 10.1155/2016/4731953.

[49] T.T Huong Le, H.Kim, T.Park, « An Effective Classification for Dos Attacks in wireless Sensor Network». Conference paper, july 2018,(p.4).South Korea.

[50] "WEKA", Master d'informatique, Apprentissage à partir d'exemples, janvier 2009.

[51] S. T. Zargar, J. Joshi and D. Tipper, "A Survey of Defense MechanismsAgainst Distributed Denial of Service (DDoS) Flooding Attacks," in IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2046-2069, FourthQuarter 2013. doi: 10.1109/SURV.2013.031413.00127