Détection intelligente de brouillage dans les réseaux 5G

par

Marouane HACHIMI

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION DE LA MAÎTRISE
AVEC MEMOIRE EN GÉNIE ÉLECTRIQUE
M. Sc. A.

MONTRÉAL, LE 03 JUIN 2020

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE:

M. Georges Kaddoum, Directeur de Mémoire
Département de génie électrique à l'École de Technologie Supérieure

M. Ghyslain Gagnon, Président du Jury
Département de génie électrique à l'École de Technologie Supérieure

M. Kim Khoa Nguyen, membre du jury
Département de génie électrique à l'École de Technologie Supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 27 MAI 2020

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Au terme de la rédaction de ce mémoire, c'est un devoir agréable d'exprimer en quelques lignes la reconnaissance que je dois à tous ceux qui ont contribué de loin ou de près à l'élaboration de ce travail, qu'ils trouvent ici mes vifs respects et ma profonde gratitude.

Je présente ma profonde reconnaissance à mon professeur et encadrant, M. Georges Kaddoum, pour son suivi et pour son énorme soutien qu'il n'a cessé de me prodiguer tout au long de la période de mon mémoire.

Je remercie tous les membres de ma famille qui m'ont apporté leur soutien et leur aide au cours de ces années d'études, et plus spécialement ma chère mère et mon cher père.

J'adresse aussi mes vifs remerciements aux membres du jury pour avoir bien voulu examiner et juger ce travail.

Je ne laisserai pas cette occasion passer, sans remercier tous mes collègues au laboratoire LACIME pour leur aide et leurs précieux conseils.

Enfin, mes remerciements à tous ceux qui ont contribué de près ou de loin au bon déroulement de ce projet.

Détection intelligente de brouillage dans les réseaux 5G

Marouane HACHIMI

RÉSUMÉ

Dans les réseaux 5G, le réseau d'accès radio de nuage (C-RAN), également appelé réseau d'accès radio centralisé, est considéré comme une architecture future prometteuse en termes de minimisation de la consommation d'énergie et d'allocation efficace des ressources. Il fournit des infrastructures de nuage en temps réel, une radio coopérative et un traitement centralisé des données. Récemment, étant donné leur vulnérabilité aux attaques malveillantes, la sécurité des réseaux C-RAN a attiré une attention considérable. Parmi les diverses techniques de détection d'intrusion basées sur les anomalies, la plus prometteuse est la détection d'intrusion basée sur l'apprentissage machine (ML-IDS) car elle apprend avec moins d'interventions humaines. Dans ce sens, de nombreuses solutions ont été proposées qui soit elles ne sont pas très précises en termes de classification d'attaques, soit elles n'offrent qu'une seule couche de détection d'attaques. Ce mémoire se concentre sur le déploiement d'un système de détection d'intrusion (IDS) à plusieurs niveaux dans l'architecture C-RAN basé sur l'apprentissage profond qui peut détecter et classer plusieurs types d'attaques de brouillage : brouillage constant, brouillage aléatoire, brouillage trompeur, et brouillage réactif. Ce déploiement assure une sécurité accrue en minimisant les faux négatifs dans la classification. L'évaluation expérimentale de la solution proposée est réalisée sur les données WSN-DS (Wireless Sensor Networks DataSet), qui est un ensemble de données de réseau sans fil dédié à la détection des intrusions. La précision de la classification finale d'attaques est de 94,51% avec un taux de faux négatifs de 7,84%.

Mots clés: Réseaux d'accès radio de nuage, 5G, attaques par brouillage, système de détection d'intrusion, apprentissage profond, apprentissage supervisé, WSN-DS.

Smart jamming detection in 5G networks

Marouane HACHIMI

ABSTRACT

In 5G networks, the Cloud Radio Access Network (C-RAN) is considered a promising future architecture in terms of minimizing energy consumption and allocating resources efficiently by providing real-time cloud infrastructures, cooperative radio, and centralized data processing. Recently, given their vulnerability to malicious attacks, the security of C-RAN networks has attracted significant attention. Among various anomaly-based intrusion detection techniques, the most promising ones are the machine learning-based intrusion detection as they learn without human assistance. In this direction, many solutions have been proposed, but they show either low accuracy in terms of attack classification or they offer just a single layer of attack detection. This research focuses on deploying a multi-stage machine learning-based intrusion detection (ML-IDS) in 5G C-RAN that can detect and classify four types of jamming attacks, namely constant jamming, random jamming, deceptive jamming, and reactive jamming. This deployment enhances security by minimizing the false negatives in C-RAN architectures. The experimental evaluation of the proposed solution is carried out using the Wireless Sensor Networks DataSet WSN-DS, which is a dedicated wireless dataset for intrusion detection. The final classification accuracy of attacks is 94.51% with a 7.84% false negative rate.

Keywords: Cloud Radio Access Network, 5G, jamming attacks, intrusion detection system, deep learning, supervised learning, WSN-DS.

TABLE DES MATIÈRES

Page

LISTE DES TABLEAUX

Page

LISTE DES FIGURES

Page

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

RAN	Radio Access Network
C-RAN	Cloud Radio Access Network
IDS	Intrusion Detection System
S-IDS	Signature-based Intrusion Detection System
A-IDS	Anomaly-based Intrusion Detection System
ML-IDS	Machine Learning based Intrusion Detection System
K-IDS	Knowledge Based Intrusion Detection System
DM-IDS	Data Mining based Intrusion Detection System
SA-IDS	Statistical Anomaly based Intrusion Detection System
WSN-DS	Wireless Sensor Networks-DataSet
SVM	Support Vector Machine
NB	Naive Bayes
DT	Decision Tree
RF	Random Forest
K-NN	K-Nearest Neighbour
LR	Logistic Regression
ML	Machine Learning
ANN	Artificial Neural Network
RF	Random Forest
IBM	International Business Machines corporation
2G	Second Generation of mobile communications networks
3G	Third Generation of mobile communications networks
4G	Fourth Generation of mobile communications networks

5G	Fifth Generation of mobile communications networks
BS	Base Station
MS	Mobile Station
RRH	Remote Radio Head
BBU	Baseband Unit
DoS	Denial of Service
TDMA	Time Division Multiple Access
IEEE	Institute of Electrical and Electronics Engineers
CPU	Central Processing Unit
RAM	Random Access Memory
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

INTRODUCTION

Le passage à la cinquième génération des réseaux de télécommunication (5G) est imminent. Il suscite un grand enthousiasme pour plusieurs acteurs dont les opérateurs de télécommunications, les concepteurs et fabricants d'équipements de télécommunications, les fournisseurs de services, les fournisseurs d'applications, les petites et moyennes entreprises, les organismes de standardisation et les organismes gouvernementaux. Ce passage fait ainsi l'objet de plusieurs et divers travaux de recherches si bien dans les milieux universitaires que dans des domaines de l'industrie. Plusieurs travaux de déploiement sont déjà entamés dans certains pays d'Amérique du Nord, d'Europe et d'Asie. Le lancement officiel de cette nouvelle génération est prévu pour 2020 (Illy, P. (2018)).

La 5G se différencie de ces prédécesseurs par plusieurs progrès techniques qui sont principalement le débit élevé (20 Gb/s), la faible latence (<1 ms), la mobilité élevée (500 km/h), la densité de connexion élevée (1 million de connections/km^2) et la faible consommation électrique (réduction de 90%).

La prochaine génération de réseaux de communications mobiles 5G a choisi le réseau d'accès radio de nuage (C-RAN) comme architecture typique pour supporter les nouveaux services et communications mobiles à l'horizon 2020.

En tant qu'architecture de réseau sans fil mobile prometteuse, comparée au RAN traditionnel, le C-RAN présente des avantages incomparables notamment une faible consommation d'énergie, une réduction du nombre de stations de base (BS) et des dépenses économiques en capital et en exploitation. Il peut également améliorer la capacité du réseau et le taux d'utilisation de la station de base.

Problématique

Récemment, la sécurité du réseau C-RAN a suscité une attention et des préoccupations particulières vu que son déploiement va impliquer une génération de données plus riches qualitativement et quantitativement. Ces réseaux seront des cibles privilégiées pour des attaquants de diverses natures (Miranda *et al.* (2020)).

Par ailleurs, la 5G est constituée d'un ensemble de changements conceptuels et technologiques importants *(network function virtualisation, network slicing)* qui transforment complètement l'architecture des réseaux *(Software defined Networks)*. Ces transformations conceptuelles, technologiques et architecturales introduisent de nouvelles exigences fondamentales en termes de sécurité au vu des nouveaux points de vulnérabilité, des nouveaux modèles de confiance, des nouveaux modèles de prestation de service et des préoccupations accrues en matière de protection de la vie privée (Illy, P. (2018)).

En plus de ces deux problèmes déjà présentés s'ajoute la sophistication croissante d'attaques qui font de plus en plus font appel à des technologies de pointe et bénéficient de plus de ressources en termes de puissance de calcul et de capacité de stockage.

Ces problématiques font donc de notre projet une étude à la foi nouvelle, importante et exigeante.

Objectif et méthodologie

L'objectif principal du projet est de proposer des solutions de détection d'intrusion pour les réseaux C-RAN de la 5G. D'une part, ces solutions doivent exploiter des approches innovatrices dans la détection d'intrusion notamment les algorithmes d'apprentissage machine et profond. D'autre part ces solutions doivent être conçues en se basant sur les nouveaux concepts, les nouvelles technologies et les nouvelles architectures introduites par la 5G. Ces solutions doivent

être en mesure de détecter non seulement les attaques standards, mais aussi les nouvelles attaques sophistiquées, tout en réduisant la latence et en maximisant la précision de détection.

Afin d'atteindre cet objectif, nous allons procéder comme suit :

- nous allons commencer par décrire l'évolution des réseaux de communication mobile ainsi que les défis qui attendent le réseau C-RAN. Ensuite, nous allons décrire les attaques qui ciblent l'architecture de ce réseau. Après cela, nous allons présenter les méthodes et les approches de détection d'intrusion les plus citées dans la littérature.
- nous allons ensuite proposer une architecture de déploiement d'un système de détection d'intrusion (IDS) basé sur l'apprentissage profond dans le réseau C-RAN qui a pour but de classifier quatre types d'attaques par brouillage. Un déploiement qui assure une sécurité accrue dans les réseaux C-RAN et une grande précision de classification d'attaques.
- finalement, en utilisant la base de données WSN-DS, qui est une base de données de réseau sans fil dédiée à la détection des intrusions, une combinaison de deux algorithmes de classification sera mise en œuvre. Nous allons aussi générer un modèle qui pourra être déployé directement dans le cœur de l'architecture du C-RAN afin de minimiser les faux négatifs.

Publication

Le travail présenté dans ce mémoire a été accepté pour publication dans la conférence "IEEE ISNCC 2020" sous le nom de "Multi-stage Jamming Attacks Detection using Deep Learning Combined with Kernelized Support Vector Machine in 5G Cloud Radio Access Networks" (Hachimi *et al.* (2020)).

Plan du mémoire

Ce mémoire est organisé comme suit : le premier chapitre définit le réseau C-RAN et son évolution jusqu'à présent. Nous détaillons aussi les différentes attaques qui ciblent cette nouvelle architecture.

Le deuxième chapitre présente les méthodes et les approches les plus citées dans la littérature, ainsi que l'approche de détection finale utilisée.

Le troisième chapitre montre l'architecture de déploiement de notre IDS basé sur l'apprentissage machine et profond (Singh *et al.* (2020)) dans le cœur du réseau C-RAN.

Le quatrième chapitre présente les résultats d'implémentation des deux algorithmes de classification ainsi qu'un tableau de comparaison des faux négatifs avant et après implémentation de notre modèle.

La figure 0.1 contient une représentation graphique du plan de ce mémoire.

Figure 0.1 Diagramme des chapitres

CHAPITRE 1

LE RÉSEAU C-RAN

1.1 Introduction

Le réseau d'accès radio est un composant du réseau mobile qui établit la connexion entre les utilisateurs mobiles et le réseau central. C'est un élément essentiel de tout réseau mobile qui permet l'établissement d'une communication efficace entre la station de base et les utilisateurs finaux. Au cours des dernières années, avec la forte croissance du nombre de clients, la consommation de trafic de données provenant de terminaux sans fil ne cesse d'augmenter (Yang *et al.* (2015)). De plus, selon une étude menée par le "BeijingKey Laboratory of Network System Architecture and Convergence, China" sur la pénétration de l'Internet mobile dans le monde de 2013 jusqu'au 2019, en 2014, juste 48,8% des téléphones mobiles de la population mondiale avaient accès à l'Internet. Ce chiffre a atteint 61,2% en 2018, avec une augmentation annuelle moyenne de 8,3% du nombre d'appareils mobiles.

Il est mentionné dans (Chih-Lin *et al.* (2014)) qu'une part importante de la consommation d'énergie des réseaux mobiles provient des réseaux d'accès radio (RAN). De plus, en raison de la pénurie de spectre et de bande passante, les RAN traditionnels ne sont pas en mesure de répondre à la demande croissante des utilisateurs mobiles. L'architecture des réseaux d'accès radio de nuage (C-RAN) pourrait être une solution pour améliorer les performances et augmenter la flexibilité afin de surmonter les problèmes des RAN traditionnels (Hadzialic *et al.* (2013)).

1.2 Contexte et motivation du projet

Les opérateurs mobiles recherchent de plus en plus les infrastructures en temps réel, la radio coopérative, le traitement centralisé des données et les réseaux d'accès radio de nuage pour répondre aux besoins des utilisateurs finaux. Depuis que IBM a défini le concept de C-RAN en 2010, cette technologie a attiré beaucoup d'attention dans le monde entier parce qu'elle a résolu

tous les défauts du RAN traditionnel. En 2020, la 5G a choisi le C-RAN comme architecture typique pour supporter ses nouveaux services et communications mobiles (Buzzi *et al.* (2016)).

1.3 3

Il y a eu un grand progrès dans la communication mobile sans fil depuis les dernières années. Une évolution qui se compose de plusieurs générations et qui se poursuit encore aujourd'hui. Cette progression de la communication mobile sans fil a commencé avec la 1G suivie de la 2G, 3G, 4G, et la 5G.

1.3.1 La deuxième génération (2G)

Comme le montre la figure 1.1, dans la deuxième génération (2G), toutes les fonctionnalités sont groupées ensemble, l'antenne reçoit le signal et elle l'envoie directement à la station de base (BS) par l'intermédiaire d'un long câble coaxial. Mais parmis les inconvénients de cette architecture c'est que plus le câble est long, plus la perte ou l'atténuation du signal est importante.

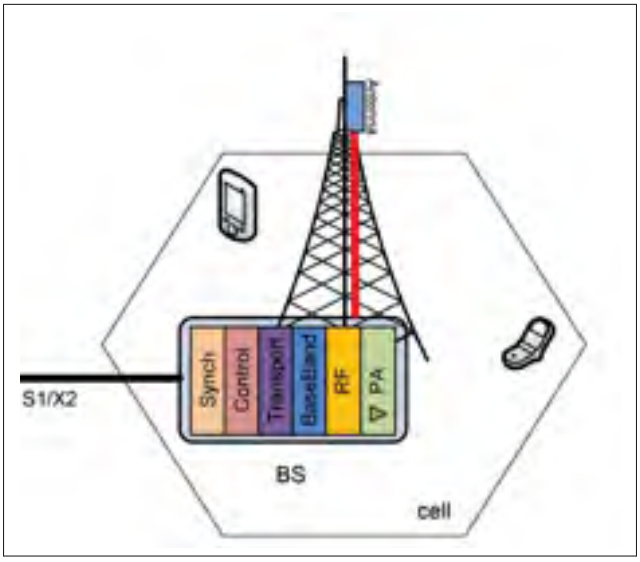

Figure 1.1 Macro station de base traditionnelle

1.3.2 La troisième génération (3G)

La troisième génération (3G) a réussi à combler cette lacune en séparant les fonctionnalités de la BS en deux parties : Remote Radio Head (RRH) et Baseband Unit (BBU). La figure 1.2 montre que chaque RRH est relié à une BBU (one-to-one) par un câble optique. Le câble optique est utilisé dans ce cas, car il présente une très faible perte et il est moins cher que le câble coaxial.

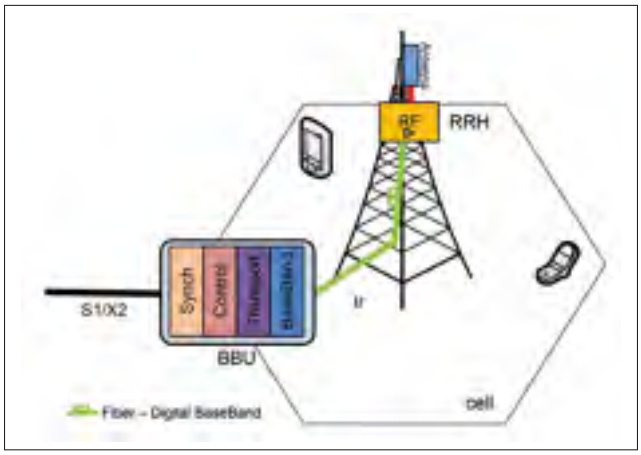

Figure 1.2 BS avec RRH (one-to-one)

1.3.3 La quatrième génération (4G)

Pour la 4G, qui a adopté le réseau d'accès radio (RAN) comme architecture, comme le montre la Figure 1.3, une BBU pourrait desservir plusieurs RRH (one-to-many) afin de réduire le coût de déploiement des BBUs.

1.3.4 La future génération (5G)

Enfin, le 5G adopte le C-RAN comme architecture typique. La figure 1.4 montre que l'architecture C-RAN regroupe toutes les BBUs dans un seul nuage qu'on appelle le BBU Pool et les RRHs sont répartis sur plusieurs sites. En conséquence, nous avons une architecture centralisée et flexible avec une faible consommation d'énergie (Boulos *et al.* (2015)).

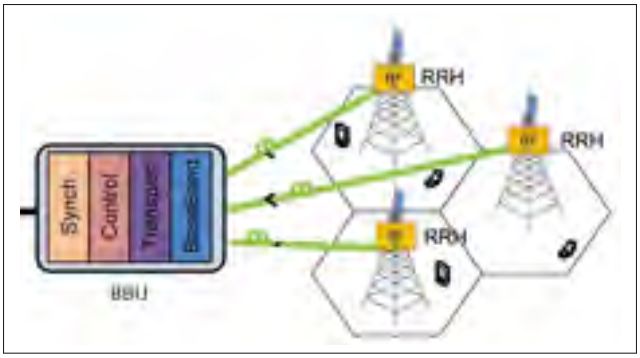

Figure 1.3 BS avec RRH (one-to-many)

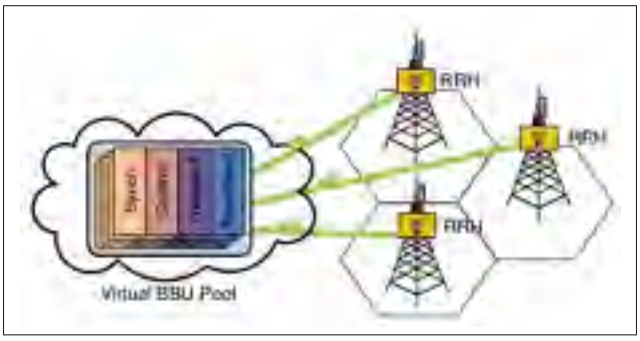

Figure 1.4 C-RAN avec RRHs

1.4 Architecture du réseau C-RAN

L'architecture C-RAN est conçue pour permettre aux opérateurs mobiles de déplacer l'unité de traitement en bande de base vers un emplacement central pour prendre en charge plusieurs RRHs. Le C-RAN offre aux opérateurs mobiles la possibilité de centraliser plusieurs BBUs en un seul endroit, soit sur un site cellulaire, soit dans un pool de BBU centralisé. Cela permet aux opérateurs télécom de simplifier la quantité d'équipement nécessaire sur chaque site cellulaire.

Comme le montre la Figure 1.5, l'architecture C-RAN utilise la fibre optique pour connecter l'équipement de la station de base aux BBU pools. Dans certaines architectures, les BBUs sont reliées entre elles et peuvent partager des informations, tandis que dans d'autres, elles sont simplement situées dans le même bâtiment. La colocalisation des BBUs est de plus en plus populaire pour les porteuses qui déploient des systèmes d'antennes distribuées.

Le déploiement d'une architecture C-RAN permet également aux opérateurs d'économiser de l'argent, car il permet de réduire les coûts d'au moins deux façons. Tout d'abord, l'immobilier est presque toujours moins cher dans un centre de données que dans une tour de téléphonie cellulaire. Grâce à cette architecture, les opérateurs mobiles peuvent consolider les équipements de stations de base pour plusieurs sites cellulaires dans un bureau central ou un centre de données. Deuxièmement, la perte de puissance est beaucoup plus faible avec la fibre optique qu'avec le câble, donc la connexion par fibre optique associée au C-RAN peut réduire les frais d'exploitation.

En ce qui concerne le déploiement d'architectures RAN en nuage, ce modèle permettra aux opérateurs mobiles d'être en mesure de répondre aux exigences croissantes en matière de latence, de débit de données et de volume de trafic grâce à l'utilisation de techniques de virtualisation des fonctions réseau et de capacités de traitement des centres de données dans leurs réseaux, qui permettent la mutualisation des ressources, l'interopérabilité des couches et l'efficacité spectrale (Moreira *et al.* (2018)).

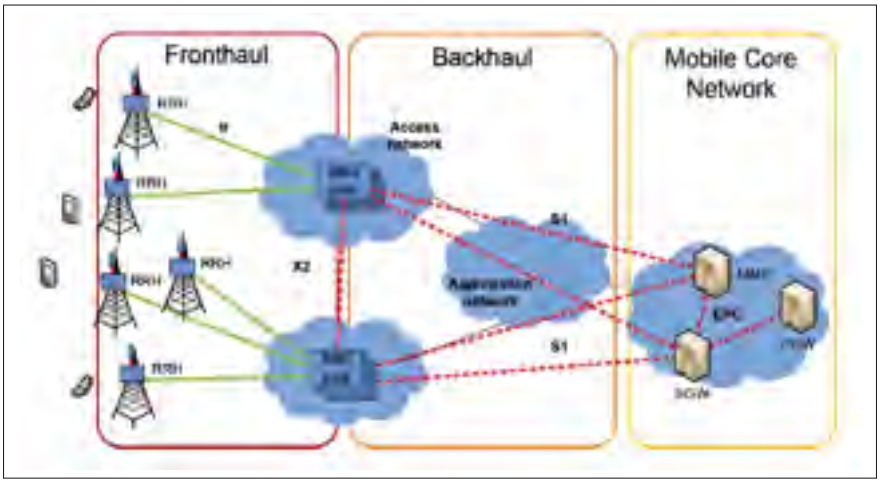

Figure 1.5 Architecture du réseau C-RAN

1.5 Avantages de l'architecture C-RAN

Une BBU centralisée offre de nombreux avantages, dont la possibilité de mettre en œuvre des technologies avancées, la virtualisation des ressources et le déploiement de services à la périphérie. Un BBU pool centralisé signifie que les BBUs sont situés au centre du réseau tandis que les RRU sont distribuées. Cela permet au C-RAN d'avoir plusieurs avantages par rapport aux réseaux cellulaires traditionnels dans lesquels les BBUs sont distribués :

- le premier avantage est la possibilité d'implémenter diverses technologies avancées qui nécessitent un traitement élevé et qui ne peuvent pas être implémentées dans les réseaux traditionnels. Comme les BBUs peuvent être situés dans des centres de données puissants et disposent d'échanges d'informations efficaces, ils peuvent effectuer des calculs étendus qui ne peuvent être effectués dans les réseaux actuels. Par conséquent, les technologies de traitement conjoint et de partage radio coopératif deviendront possibles grâce à l'architecture C-RAN.
- de plus, avec des multiples BBUs qui partagent un pool commun, contrairement aux réseaux traditionnels le partage des ressources peut devenir faisable et donc l'allocation peut être plus flexible et sur demande. Cela peut améliorer l'utilisation des ressources, réduire la consommation d'énergie et accroître la satisfaction des utilisateurs en raison de l'augmentation du bassin de ressources.
- de plus, avec des réseaux aussi étendus et distribués, les services peuvent être déployés à la périphérie du réseau plutôt qu'au cœur. Comme les serveurs C-RAN sont puissants et ont de grandes capacités de calcul, la réalisation du déploiement de services de périphérie devient beaucoup plus facile. De cette façon, les services se rapprocheront de l'utilisateur et donc les réponses sont plus rapides déclenchant une meilleure satisfaction chez les utilisateurs. De plus, cela peut réduire la charge dans les réseaux de liaison terrestre ce qui les rendrait plus flexibles et évolutifs que leur état actuel.
- par conséquent, la réalisation d'une BBU centralisée dans le C-RAN peut présenter de nombreux avantages, y compris des capacités de traitement et de partage des ressources

communes, en plus de la mise en œuvre de services à la périphérie. Cela peut accroître la satisfaction des utilisateurs, améliorer l'utilisation des ressources et réduire la pression sur le serveur central.

1.6 Défis du C-RAN

Des principaux défis confrentent l'évaluation du C-RAN, ces derniers comprennent les capacités élevées de Fronthaul nécessaires, la coopération et le regroupement des BBUs, les techniques de virtualisation utilisées, la sécurité et bien d'autres encore. Les plus importantes d'entre elles seront examinées ci-dessous.

1.6.1 Capacités du fronthaul nécessaires

La liaison Fronthaul entre les BBUs et les RRU doit être dotée d'une grande capacité de bande passante avec des exigences de délais et de coûts réduits. Une architecture qui nécessite un énorme surcoût de communication sur la liaison frontale. Par conséquent, une bande passante élevée est requise au niveau du fronthaul, ce qui ne peut pas être atteint par la communication sans fil.

La communication par fibre optique peut donner la bande passante élevée requise pour résoudre ce problème. Cependant, la fibre optique a un problème de coût très élevé qui n'est pas à la portée de la plupart des fournisseurs de services cellulaires. Par conséquent, une solution adéquate prenant en compte le retard, la bande passante et le coût doit être prise en compte dans de tels systèmes avant qu'ils ne deviennent une réalité.

1.6.2 Coopération entre les BBUs

Les BBUs d'un même pool doivent coopérer afin de soutenir le partage des données, la programmation et la collecte des retours d'expérience des utilisateurs. Une telle coopération n'est pas définie et pose un défi quant à la façon de gérer la vie privée des utilisateurs, la bande passante élevée et la faible latence de communication entre ces BBUs.

1.6.3 Technologie de virtualisation

Les techniques de virtualisation favorisent le traitement distribué et le partage des ressources entre plusieurs BBUs, ce qui représente un autre défi pour le C-RAN. Le traitement doit être en temps réel et dynamique afin de supporter les charges cellulaires changeantes. En outre, l'exigence relative aux nuages sur lesquels les BBUs seront implémentés sera différente des exigences connues en matière de nuages informatiques. Ais l'infrastructure de nuage doit être adaptée pour répondre à ces exigences. Par conséquent, la virtualisation est un autre défi critique qui affecte le déploiement actuel du C-RAN dans la pratique.

1.6.4 Sécurité et confidentialité

Un autre défi important du C-RAN est celui de la sécurité en termes de confidentialité des utilisateurs et des personnes de confiance. Comme les ressources sont partagées entre les BBUs, la violation de la vie privée des utilisateurs et l'accès à des données présumées sécurisées est une possibilité, en particulier dans une telle architecture distribuée. En outre, les parties sont censées être dignes de confiance dans les réseaux C-RAN, y compris les BBUs et les RRHs. De telles hypothèses pourraient être invalides, surtout compte tenu du grand nombre d'utilisateurs abonnés à de tels systèmes. Un utilisateur malveillant peut profiter d'un système virtualisé de cette taille pour mal se comporter et menacer le système. En plus des vulnérabilités existent dans les systèmes cellulaires traditionnels, le réseau C-RAN présenterait un autre défi en matière de sécurité qui n'était pas envisagé ou qui l'était moins auparavant.

1.7 Attaques sur le réseau C-RAN

Récemment, en raison de leur vulnérabilité aux attaques malveillantes, la sécurité des réseaux C-RAN a suscité une attention et des préoccupations particulières. En raison de la nature ouverte des réseaux sans fil, les utilisateurs autorisés et illégitimes peuvent accéder au canal de communication (Mavoungou *et al.* (2016)). Ainsi, le C-RAN hérite de toutes les attaques qui peuvent être effectuées dans un réseau sans fil.

1.7.1 Attaque d'écoute clandestine

Une attaque d'écoute clandestine, également connue sous le nom d'attaque de reniflement ou d'espionnage, est une incursion où quelqu'un tente de voler des informations que des ordinateurs, des téléphones intelligents ou d'autres dispositifs transmettent sur un réseau. Comme le montre la figure 1.6 Tirée de Ioannis *et al.* (2016) et le site (https ://dl.acm.org/doi/10.1145/3003733.3003791), une attaque d'écoute clandestine profite des communications réseau non sécurisées pour accéder aux données envoyées et reçues. Les attaques d'écoute sont difficiles à détecter parce qu'elles ne donnent pas l'impression que les transmissions réseau fonctionnent anormalement (Atallah *et al.* (2019)).

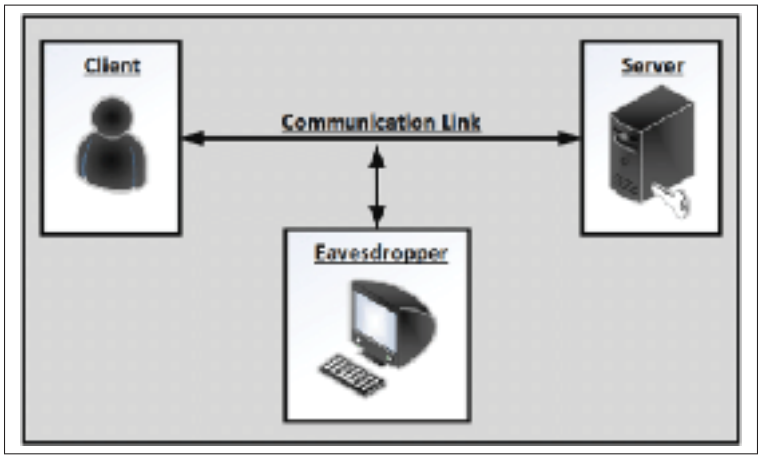

Figure 1.6 Attaque d'écoute clandestine

1.7.2 Attaque d'émulation d'utilisateur primaire (PUEA)

Les attaques d'émulation d'utilisateur primaire (PUEA) sont un type d'attaques faciles à lancer, mais difficiles à détecter dans les réseaux sans fil. Ces attaques imitent les signaux d'un utilisateur primaire (PU) afin d'occuper égoïstement les ressources du spectre ou de mener d'attaques de déni de service (DoS) (Deepa *et al.* (2013)).

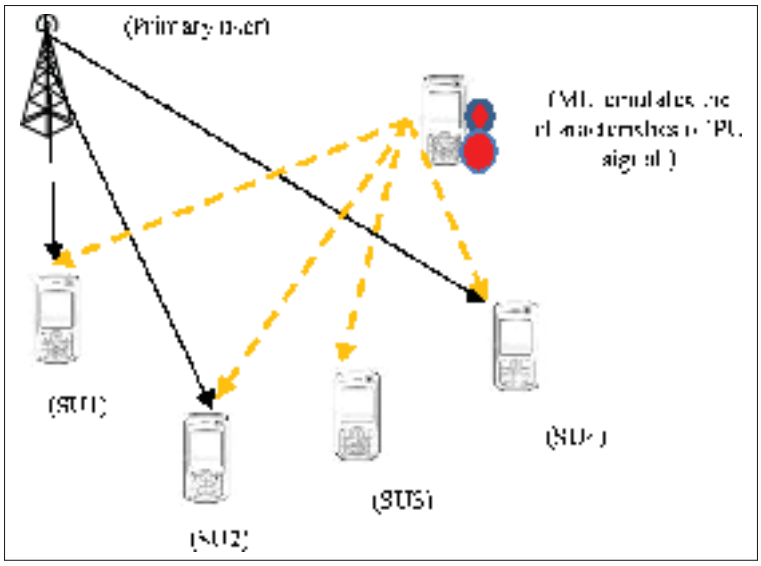

Figure 1.7 Attaque d'émulation d'utilisateur primaire (PUEA)

1.7.3 Attaque d'usurpation d'identité

Une attaque d'usurpation d'identité est une attaque dans laquelle un adversaire assume avec succès l'identité de l'une des parties légitimes dans un système ou dans un protocole de communication comme le montre la figure 1.8 Tirée de (http ://slideplayer.com/slide/4128699).

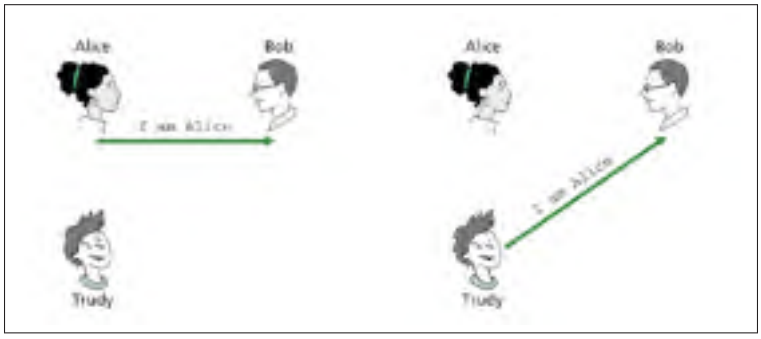

Figure 1.8 Attaque d'usurpation d'identité

1.7.4 Attaque de brouillage

Les attaques de brouillage représentent la menace de sécurité la plus grave dans le domaine des communications sans fil, car elles peuvent facilement empêcher les autres nœuds d'utiliser le canal pour communiquer en occupant le canal sur lequel ils communiquent (Kong *et al.* (2016)).

Un brouilleur est sûrement utilisé dans cette attaque. Comme le montre la Figure 1.9, c'est un équipement qui peut perturber le signal d'un nœud en augmentant sa densité spectrale de puissance (PSD). Il existe plusieurs types de brouilleurs qui peuvent être utilisés contre les réseaux C-RAN, à savoir : brouilleur constant, brouilleur aléatoire, brouilleur trompeur et brouilleur réactif (Jameel *et al.* (2019)).

1.7.4.1 Brouilleur constant

Un brouilleur constant produit en continu un bruit de forte puissance sous forme de bits aléatoires. Il fonctionne indépendamment de la détection du trafic sur le canal.

1.7.4.2 Brouilleur aléatoire

Un brouilleur aléatoire fonctionne de façon aléatoire dans deux états : sommeil et brouillage. Pendant l'état de veille, il est inactif et pendant l'état de brouillage, il agit comme un brouilleur constant.

1.7.4.3 Brouilleur trompeur

C'est un brouilleur constant qui transmet continuellement des paquets réguliers. Il est plus difficile à détecter, car il transmet des paquets légitimes au lieu de bits aléatoires (Atallah *et al.* (2015)).

1.7.4.4 Brouilleur réactif

Un brouilleur réactif ou intelligent est activé dès qu'il détecte la transmission sur le canal et commence à envoyer des paquets illégitimes. Si le canal est inactif, il reste inactif et continue à le détecter (Atallah *et al.* (2015)).

1.8 Conclusion

Ce chapitre a passé en revue certains des principes de base, des avantages et des défis actuels de la technologie C-RAN qui prend en charge la virtualisation des stations de base dans les futurs réseaux cellulaires. Cette technologie possède une architecture centralisée composée d'un BBU pool centralisé, RRU, et d'une liaison fronthaul pour les connecter. La virtualisation des stations de base présente de nombreux avantages en termes de coût et de consommation énergétique.

Toutefois, en raison de la nature ouverte du réseau sans fil C-RAN, l'aspect de la sécurité reste un défi très important vu qu'elle a suscité une attention et des préoccupations particulières récemment. De plus, les attaques de brouillage représentent la menace de sécurité la plus sérieuse dans le domaine des communications sans fil. Plusieurs formes de cette attaque peuvent être utilisées contre les réseaux C-RAN, à savoir : le brouillage constant, le brouillage aléatoire, le brouillage trompeur et le brouillage réactif. L'objectif est de mettre en place un IDS (Intrusion Detection System) qui aura pour fonction de détecter ces quatre types d'attaques de brouillage.

CHAPITRE 2

SYSTÈME DE DÉTECTION D'INTRUSION

2.1 Introduction

Ce chapitre est composé de trois sections. Dans la première section, nous introduisons quelques définitions sur les systèmes de détection d'intrusion. Dans la deuxième section, nous présentons les types de systèmes de détection d'intrusion qui existent. Nous détaillons, dans la troisième section, les différentes méthodes et approches de détection d'intrusion ainsi que l'approche de détection que nous allons utiliser par la suite.

2.2 Définition du système de détection d'intrusion (IDS)

Quel que soit le niveau des moyens techniques mis en place pour la prévention d'attaques dans un système informatique, il peut succomber face à un adversaire endurant, utilisant des techniques plus évoluées ou des moyens plus avancés (Moreira *et al.* (2018)). De ce fait, il faudra qu'en dessous de toute couche de prévention, qu'il ait une couche de détection d'intrusion. C'est le fondement du développement des systèmes de détection d'intrusion, plus connus sous le nom d'IDS (Intrusion Detection Systems) (Illy, P. (2018)).

Un IDS est un système de sécurité qui surveille un système informatique et analyse le trafic pour déceler des actions qui tendent à compromettre les objectifs de sécurité (confidentialité, intégrité, disponibilité…) définis pour ce système informatique.

Le processus d'IDS est décrit dans la Figure 2.1. Les données circulant dans le réseau sont extraites par un "préprocesseur de données". Ensuite, le moteur de détection analyse ces derniers et, en se basant sur les modèles de détection, il détermine s'il y a une intrusion ou non. Dans le cas d'intrusion, l'IDS produit une alerte et l'envoie au moteur de décision qui décide également s'il va prendre des mesures contre l'attaque ou informer l'administrateur.

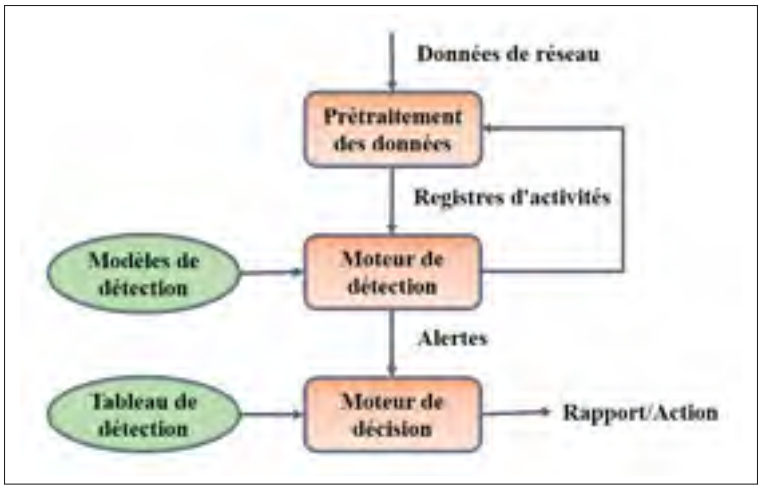

Figure 2.1 Modèle générique d'un IDS

Les recherches sur la détection d'intrusion ont commencé il y a près de 40 ans. En 1980, James P.Anderson, membre du "Defense Science Board Task Force on Computer Security" de l'U.S Air Force, publie "Computer Security Threat Monitoring and Surveillance", un rapport souvent reconnu comme ayant introduit l'IDS automatisé. Actuellement, les recherches dans le domaine sont encore plus dynamiques, notamment en raison des évolutions technologiques comme le cloud computing, le big data et l'intelligence artificielle.

2.3 Les types d'IDS

Au regard de la diversité d'attaques qui peuvent être mises en oeuvre, la détection d'intrusion doit se faire à plusieurs niveaux (Garg *et al.* (2020)). Il existe différents types d'IDS selon l'endroit qu'ils surveillent et ce qu'ils contrôlent ou selon leurs fonctions (Illy, P. (2018)).

2.3.1 Les IDS réseaux

Ils sont connus aussi sous le nom de Network-based IDS. Les IDS réseaux analysent et interprètent les paquets circulant sur un réseau (ou un segment du réseau) afin de repérer les paquets à contenus malicieux. Le paquet est analysé sur toutes ses couches (réseau, transport, applica-

tion). Par dissection des paquets et la connaissance des protocoles, les NIDS sont capables de détecter des paquets malveillants conçus pour outrepasser un pare-feu.

Ce type d'IDS a l'avantage d'être plus facile à protéger (contre les attaques sur l'IDS lui-même) du faîte qu'ils ne font qu'une observation du trafic. Cependant, une des contraintes des NIDS est que, pour pouvoir écouter l'ensemble des paquets, ils nécessitent une bande passante qui est proportionnelle à l'importance du trafic. Aussi, le positionnement des NIDS dans le réseau doit être stratégique afin de pouvoir surveiller tout le trafic. Par ailleurs les NIDS présentent des limites dans la protection des réseaux aux trafics chiffrés.

Quelques exemples de NIDS sur le marché sont NetRanger, NFR, Snort, DTK et ISS RealSecure.

2.3.2 Les IDS hôtes

Les systèmes de détection d'intrusion basés sur l'hôte, aussi appelés Host-based IDS, analysent exclusivement les activités concernant l'hôte sur lequel ils sont installés (serveur, poste client, pare-feu, etc.), recherchant des activités suspectes. La détection peut se faire en utilisant les logs d'audit de sécurité, les logs système, le trafic réseau de l'hôte, les processus en cours d'exécutions, les accès aux fichiers, les changements de configurations des applications, etc. Le plus souvent les HIDS sont déployés sur les hôtes critiques comme les serveurs contenant des informations de sensibilités élevées et les serveurs publiquement accessibles.

Étant focalisés sur la sécurité d'un seul hôte, les HIDS ont l'avantage d'avoir plus de précision sur les variétés d'attaques. Aussi l'impact d'une attaque peut être constaté et permet une meilleure réaction. d'attaques dans un trafic chiffré peuvent être détectées (impossible avec un IDS réseau). Cependant les HIDS sont plus vulnérables aux attaques de dénis de service. Aussi, en raison de leurs volumes de données, l'analyse des logs peut nécessiter d'importantes ressources (puissance de calcul et stockage).

Des exemples de HIDS sont OSSEC (Open Source Security), Tripwire, Radmind, EMERALD's eXpert-BSM AIDE (Advanced Intrusion Detection Environment) et PortSentry.

2.3.3 Les IDS hybrides

La nouvelle tendance en matière de détection d'intrusion est de combiner les NIDS et les HIDS pour concevoir des IDS hybrides. Les systèmes hybrides de détection d'intrusion sont flexibles et augmentent le niveau de sécurité. Ils combinent plusieurs localisations des systèmes IDS et recherchent si bien les attaques visant des éléments particuliers que celles visant l'ensemble du système. Un exemple d'IDS hybride est l'ISS RealSecure.

2.3.4 Les IDS basés sur une application

Les IDS basés sur les applications (ABIDS pour Application-Based IDS) sont un sous-groupe des IDS hôtes, parfois mentionnés séparément. Ils contrôlent l'interaction entre un utilisateur et un programme en ajoutant des fichiers de logs afin de fournir de plus amples informations sur les activités. Opérant entre utilisateur et programme, il est facile pour l'ABIDS de filtrer tout comportement notable. Ses principaux avantages sont de travailler en clair (contrairement aux NIDS, par exemple) d'où une analyse plus facile, et la possibilité de détecter et d'empêcher des commandes particulières dont l'utilisateur pourrait se servir avec le programme. Deux inconvénients majeurs sont identifiés : le peu de chances de détecter, par exemple, un cheval de Troie (puisque l'ABIDS n'agit pas dans l'espace du noyau) ; en outre, les fichiers de logs générés par ce type d'IDS sont des cibles faciles pour les attaquants (ils ne sont pas aussi sûrs que les traces d'audit du système, par exemple).

2.3.5 Les Pots de miel

Possédant de nombreuses similitudes avec les IDS, les pots de miel sont quelquefois considérés comme faisant partie des IDS. Un pot de miel est un outil informatique (système, serveur, programme, etc.), volontairement exposé et vulnérable à une ou plusieurs failles connues, des-

tiné à attirer et à piéger les pirates, tout en permettant d'observer leur comportement en pleine action et d'enregistrer leurs méthodes d'attaque pour mieux les étudier, les comprendre et les anticiper. La tâche principale d'un pot de miel consiste à analyser le trafic, c'est-à-dire à informer du démarrage de certains processus, de la modification de fichiers, etc. Permettant ainsi de créer un profil élaboré des attaquants potentiels. Tout l'art du pot de miel consiste à remonter jusqu'à l'origine de l'attaque, sans que le pirate s'en doute et que jamais il ne puisse soupçonner le fait que le site visité ne soit qu'un leurre (montage). Le réseau interne n'est pas exposé puisque le pot de miel est généralement placé dans la DMZ (zone démilitarisée).

2.3.6 Les systèmes capitonnés

Ces systèmes fonctionnent généralement avec l'un des systèmes présentés précédemment. Si l'IDS utilisé informe d'une attaque, l'attaquant est dirigé vers un hôte "capitonné". Une fois dans cet hôte, il ne peut plus causer de dommages puisque l'environnement est simulé. Cet environnement doit être aussi réaliste que possible, autrement dit, l'attaquant doit penser que son attaque a été couronnée de succès. Il est alors possible d'enregistrer, de suivre et d'analyser toute activité de l'attaquant. Les systèmes "capitonnés" présentent des inconvénients : d'une part leur usage peut être illégal (comme pour les pots de miel, ils sont tolérés en Europe) ; d'autre part, la mise en oeuvre d'un tel système est assez difficile et réclame des compétences puisque l'ensemble de l'environnement doit être simulé correctement (si l'administrateur fait une petite erreur, ce système peut être à l'origine de nouveaux trous de sécurité).

2.4 Les méthodes de détection d'intrusion

L'idée de base de la détection d'intrusion part de l'hypothèse que les comportements d'une activité intrusive sont plus ou moins différents de ceux des activités normales et sont donc détectables (Denning *et al.* (1987)). Plusieurs approches de détection d'intrusion ont été proposées dans la littérature. Généralement ces approches sont classées en quatre catégories : la détection d'anomalie, la détection basée sur les signatures et la détection basée sur des spéci-

fications et l'analyse des protocoles avec Etat. Des solutions combinant plusieurs méthodes de détection sont appelées détections hybrides.

2.4.1 La détection basée sur les signatures (S-IDS)

La détection basée sur les signatures est aussi connue sous le nom de détection d'abus (misuse detection). Cette approche vise à coder les connaissances sur les modèles de flux de données qui correspondent à des procédures intrusives sous la forme de signatures spécifiques. Ainsi, une signature est un modèle qui correspond à une menace spécifique étudiée. Les intrusions sont détectées en faisant un matching entre les évènements du système et les signatures. Les correspondances trouvées sont considérées comme des intrusions. La figure 2.2. illustre la détection basée sur les signatures Tirée de Illy, P. (2018).

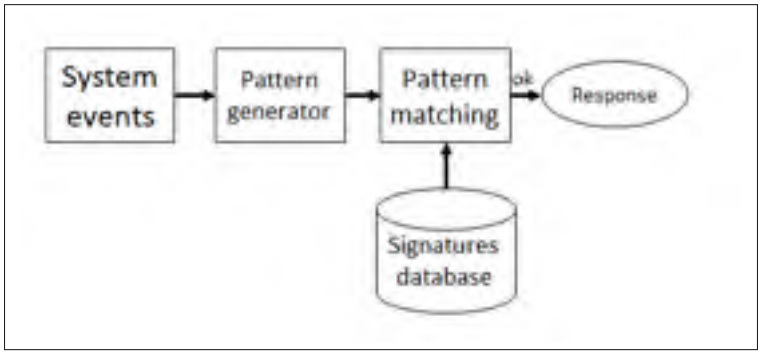

Figure 2.2 Système de détection basée sur les signatures

Le modèle comprend cinq composantes : System events qui collecte les évènements courants du système, Pattern generator qui génère les signatures des évènements du système à partir des évènements collectés, Pattern matching qui compare les signatures générées à partir des évènements courants avec celles d'attaques connues, Signatures database qui est la base de données des signatures connues. La dernière composante, response, est la réaction effectuée quand un matching est positif.

Plusieurs catégories de techniques sont couramment utilisées pour mettre en oeuvre la détection basée sur les signatures, à savoir le pattern matching, les techniques basées sur des règles, les

techniques basées sur des états et le data mining. Par contre, S-IDS a un inconvénient majeur puisque cette méthode ne parvient pas à identifier de nouvelles attaques.

2.4.2 La détection basée sur les anomalies (A-IDS)

Le principe de base de la détection d'anomalie est de modéliser durant une première période, dite phase d'apprentissage, le comportement « normal » du système en définissant une ligne de conduite (dite Baseline ou profil) (Garg, S. (2019)). Ensuite, en une seconde phase, période de détection, il est considéré comme suspect tout comportement inhabituel c'est-à-dire les déviations significatives par rapport au modèle de comportement « normal ». (Ghorbani *et al.* (2009)) propose un modèle typique de détection d'anomalie illustré par la figure 2.3.

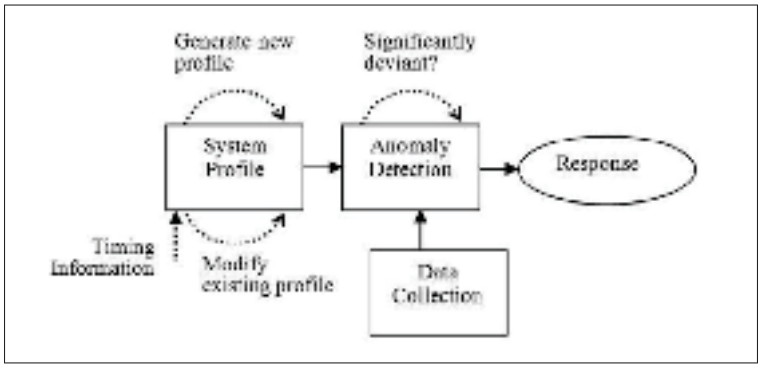

Figure 2.3 Système typique de détection basée sur les anomalies

Le système illustré ici est constitué de quatre composantes à savoir la collecte de données (Data Collection), le profil normal du système (System Profile), la détection d'anomalie (Anomaly Detection) et la réaction (response). Les activités normales du système ou les données relatives au trafic sont enregistrées par la composante de collection de données. Des techniques de modélisation spécifiques sont utilisées pour créer les profils normaux du système. La composante de détection d'anomalie détermine combien les activités en cours s'écartent des profils normaux du système et à quel seuil d'écart ces activités devraient être signalées comme anormales. Enfin, la composante de réaction signale l'intrusion et éventuellement les informations de timing correspondantes (Garg, S. (2019)).

L'avantage principal de la détection d'anomalie est sa capacité à trouver de nouvelles attaques. Ce qui constitue la plus grande limitation de la détection d'abus. Cependant, en raison des hypothèses sous-jacentes aux mécanismes de détection des anomalies, leurs taux de fausses alarmes sont en général très élevés.

De nombreuses techniques de détection anomalie ont été proposées dans la littérature. Ces modèles vont des modèles statistiques avancés à des modèles d'intelligence artificielle et des modèles biologiques basés sur les systèmes immunitaires humains. Bien qu'il soit difficile de classer ces techniques, nous pouvons les diviser en cinq catégories sur la base des enquêtes précédentes sur les systèmes de détection d'anomalie. Il s'agit notamment de modèles statistiques avancés, de modèles fondés sur des règles, de modèles d'apprentissage, de modèles biologiques et de modèles fondés sur des techniques de traitement du signal (Illy, P. (2018)).

Comme mentionné dans la figure 2.4 Tirée de Veeramreddy *et al.* (2011), il y a beaucoup des techniques de détection basée sur les anomalies telles que l'approche basée sur l'analyse statistique, l'approche basée sur la fouille de données, l'approche basée sur la connaissance et l'approche basée sur l'apprentissage machine qu'on va les détailler ci-dessous.

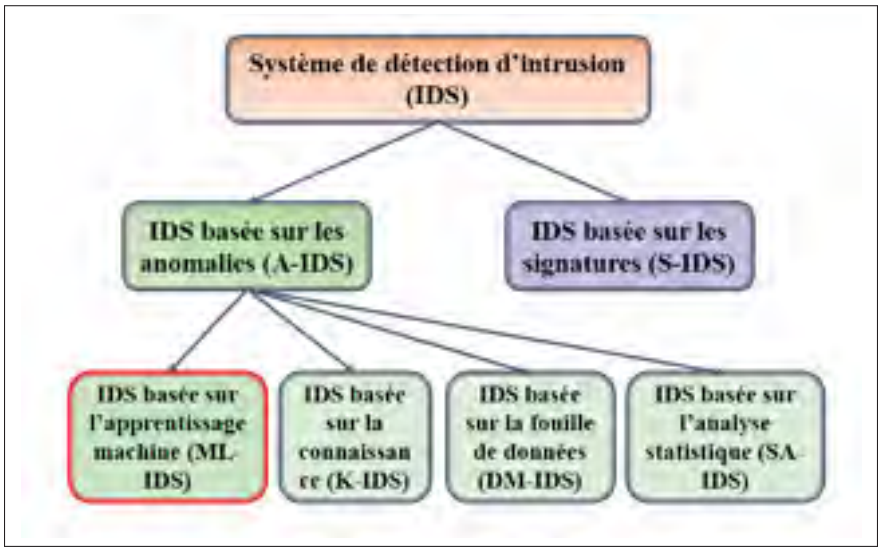

Figure 2.4 Classification hiérarchique d'un IDS

2.4.2.1 Approche basée sur l'analyse statistique

La modélisation statistique est une des techniques les plus utilisées récemment pour la détection d'intrusion. Les techniques statistiques de détection d'anomalie utilisent des propriétés et des tests statistiques pour savoir si le comportement observé dévie considérablement du comportement attendu (Qayyum *et al.* (2005)). Il y a deux étapes principales dans le processus de cette approche : la première crée des profils de comportement pour les activités normales et les activités courantes. Ces profils sont liés à travers plusieurs techniques pour détecter toute sorte de déviation par rapport au comportement normal. Les approches statistiques peuvent être classifiées en deux catégories :

Le modèle d'approche de Markov (Markov Model Approach) :

Le modèle Markovien est utilisé avec la métrique de comptage d'évènement pour déterminer la régularité d'évènements particuliers, sur la base de l'évènement qui le précède. Ce modèle caractérise chaque observation comme un état spécifique et utilise une matrice de transition d'état pour déterminer si la probabilité de l'évènement est élevée ou normale au regard des évènements précédents. Cela est très efficace quand les séquences d'activités sont particulièrement importantes. La chaîne de Markov maintien à jour sa recherche d'intrusion en examinant le système à des intervalles de temps fixes et garde les enregistrements de son état. Si un changement d'état a lieu, il calcule la probabilité pour cet état à un l'intervalle de temps donné et si cette probabilité est faible à cet intervalle de temps, alors cet événement est considéré comme une anomalie (Qayyum *et al.* (2005)).

Le modèle d'approche opérationnelle (Operational Model Approach) :

Ce modèle se base sur hypothèse et fait l'identification d'anomalie en comparant l'observation avec une limite prédéfinie. Sur la base de la cardinalité d'une observation faite sur une certaine période de temps, une alarme est générée. Le modèle opérationnel s'applique surtout aux métriques pour lesquels l'expérience a montré que certaines valeurs sont souvent associées à des

intrusions. C'est le cas par exemple d'un compteur d'événements pour le nombre d'erreurs de mot de passe pendant une brève période.

2.4.2.2 Approche basée sur la fouille de données

La fouille de données peut être une bonne solution pour la détection d'intrusion. Elle permet, à partir d'un grand volume de données, d'extraire des modèles utiles qui étaient méconnus. L'approche de fouille de données est susceptible de réduire la quantité de données qui doit être retenue pour l'étude de l'activité du système, créant ainsi des données plus pertinentes pour la détection des intrusions. Cette approche a l'avantage de pouvoir identifier les générateurs de fausses alarmes et les "mauvaises" signatures et de trouver les activités anormales qui sont d'attaques réelles. Les approches de détection basées sur la fouille de données peuvent être classées selon les techniques suivantes :

L'approche de Clustering :

L'analyse de cluster ou Clustering est une opération de regroupement d'un ensemble d'objets de telle sorte que les éléments d'un même groupe (appelé cluster) sont plus similaires (dans un sens ou un autre) les uns aux autres que les éléments d'autres clusters. Le Clustering est une technique non supervisée pour la découverte des structures dans des données non labélisées de grandes dimensions. Généralement le k-mean Clustering est utilisé pour trouver le regroupement naturel d'instances similaires. Les éléments qui se trouvent à une grande distance de tous les groupes indiquent une activité inhabituelle qui peut être considérée comme une nouvelle attaque (Ian *et al.* (2004)).

L'approche de Classification :

L'objectif principal de la classification est de se baser sur des classes déjà connues de données d'apprentissage, pour prédire les classes de nouvelles données. Les avantages des techniques de détection basées sur la classification [16], en particulier les techniques multi classes, sont qu'elles utilisent des algorithmes puissants qui permettent de distinguer les éléments apparte-

nant à différentes classes. La phase de prédiction des techniques basées sur la classification est rapide. Chaque activité du système est comparée avec le modèle précalculé.

2.4.2.3 Approche basée sur la connaissance

Dans cette approche, une collection de connaissances sur d'attaques spécifiques et les vulnérabilités des systèmes est d'abord effectuée. Ensuite on applique ces connaissances pour reconnaître les profils d'attaques et des vulnérabilités du système à sécuriser. Tout autre événement que le système est incapable de reconnaître est accepté et, par conséquent, l'exactitude de l'IDS basé sur la connaissance est considérée comme élevée. Cependant, leur principale exigence est que la base de connaissances soit suffisamment fournie et mise à jour régulièrement. Les techniques de détection basées sur la connaissance peuvent être classifiées comme :

L'approche de système expert (Expert System Approach) :

Elles sont principalement utilisées par les IDS basés sur la connaissance. Pour décrire les attaques, il y a un ensemble de règles. Les événements d'audit sont ensuite traduits en faits porteurs de leur signification sémantique dans le système expert, et ces règles et faits serviront au moteur d'inférence pour tirer des conclusions. Cette méthode augmente le niveau d'abstraction des données d'audit en y attachant une sémantique (Varun *et al.* (2009)).

L'analyse des transitions d'états (State Transition Analysis) :

Cette technique est conceptuellement identique au raisonnement basé sur un modèle. Elle définit les attaques avec un ensemble d'objectifs et de transitions et les représente sous la forme d'un diagramme de transition d'état. Le diagramme de transition d'état est une représentation graphique des actions effectuées par un intrus pour la compromission d'un système. Dans cette technique, une intrusion est considérée comme une séquence d'actions effectuées par un intrus, c'est-à-dire des pointeurs allant d'un état initial d'un à un état final qui est la compromission visée. Les diagrammes d'analyse de transition d'état reconnaissent les exigences et le compro-

mis de la pénétration. Ils dressent des listes d'actions clés qui doivent se produire pour qu'une intrusion puisse s'effectuer (Varun *et al.* (2009)).

2.4.2.4 Approche basée sur l'apprentissage machine

L'apprentissage machine est un ensemble de techniques permettant à une application d'apprendre et d'améliorer ses performances au fil de l'expérience. Il consiste principalement à la mise en place d'un système d'amélioration de la performance sur la base des résultats précédents. Cela se fait par l'auto adaptation continuelle des données d'exécution du programme sur la base des nouvelles acquisitions. Cette caractéristique rend cette technique utile dans diverses situations, mais l'inconvénient est leur mise en oeuvre coûteuse en ressource. Dans de nombreux cas, la technique d'apprentissage machine coïncide avec les techniques statistiques et celles de fouille de données (Garcia-Teodoro *et al.* (2009)).

2.4.3 L'approche de détection utilisée

Parmi les diverses techniques d'A-IDS telles que l'approche basée sur l'analyse statistique, l'approche basée sur la fouillede données, l'approche basée sur la connaissance et l'approche basée sur l'apprentissage machine, la plus prometteuse qui montre un grand potentiel est celle de l'apprentissage machine car elle peut améliorer progressivement ses performances par un auto-apprentissage sur une durée donnée et en effectuant des tâches déterminées. Dans ce sens, différentes solutions ont été proposées pour faire face aux menaces à la sécurité (Jameel *et al.* (2019)).

2.4.4 État de l'art

Syed *et al.* (2018) a proposé un nouveau système de détection d'intrusion basé sur un réseau de modulation radio pour bloquer les attaques, appelé LIDS. Ils ont mis en œuvre deux algorithmes LIDS basés sur la divergence de Kullback Leibler (KLD) et la distance de Hamming (HD). Les taux de détection obtenus sont de 98% et 88%, respectivement, avec un taux de faux

positifs de 5%. Imen *et al.* (2016) a conçu un mécanisme de détection d'intrusion pour limiter les attaques par déni de service dans les réseaux de capteurs sans fil. Elle a mis en œuvre cinq algorithmes d'apprentissage machine pour détecter et classer les attaques DoS. Oscar *et al.* (2014) a présenté une approche de détection du brouillage basée sur l'apprentissage machine, capable de détecter des brouilleurs constants et réactifs dans divers scénarios dans les réseaux 802.11. Yi *et al.* (2018) a présenté une méthode d'apprentissage machine pour lancer d'attaques de brouillage dans les communications sans fil et a également présenté une stratégie de défense. Pour les réseaux ad hoc de véhicules (VANETs). Dimitrios *et al.* (2018) a présenté une méthode de détection et de regroupement d'attaques de brouillage des radiofréquences (RF) basée sur l'utilisation de l'apprentissage machine non supervisé. Thi *et al.* (2018) a conçu un IDS basé sur l'apprentissage machine pour classer quatre attaques DoS dans les réseaux de capteurs sans fil (WSN).

2.5 Conclusion

Ce chapitre a présenté les systèmes de détection d'intrusion, ainsi que les types d'IDS qui existent. On a présenté aussi les différentes méthodes de détection d'intrusion.

Par la suite, nous avons présenté les différentes méthodes de détection. Dernièrement, l'approche de détection utilisée a été présentée en se basant sur les objectifs et les contraintes de notre projet.

CHAPITRE 3

DÉVELOPPEMENT D'UN MODÈLE DE CLASSIFICATION D'ATTAQUES DE BROUILLAGE

3.1 Introduction

Ce chapitre présente la solution proposée en exposant d'abord l'architecture de déploiement, puis en justifiant les classificateurs implémentés.

3.1.1 Aperçu du protocole LEACH

LEACH (Low Energy Aware Cluster Hierarchy) est un protocole de regroupement adaptatif et auto-organisé dans les WSNs qui se caractérise par sa simplicité et sa faible énergie. LEACH suppose que la station de base (BS) est fixe et située loin des nœuds de capteurs. L'idée principale du protocole LEACH est d'organiser les nœuds en groupes pour distribuer l'énergie entre tous les nœuds du réseau. De plus, dans chaque cluster, il y a un nœud de tête de groupe (CH) qui collecte les données reçues des capteurs de son cluster et les transmet à la station de base.

La figure 3.1 Tirée de Singh *et al.* (2017) montre la structure des nœuds du protocole de routage LEACH. Chaque cycle du protocole LEACH se compose principalement de deux phases : la phase de mise en place et la phase d'équilibre. Dans la phase de mise en place, des groupes sont formés, tandis que dans la phase d'équilibre, les données détectées sont transférées vers le nœud de l'évier (sink node).

3.1.2 L'architecture de déploiement

La Figure 3.2 ci-dessous montre la nouvelle architecture de déploiement de notre ML-IDS dans l'architecture C-RAN avec sa topologie du maillage. À gauche, nous avons une collection des CHs qui ont un accès direct aux stations de base (BS), au sommet des BS nous avons un certain nombre d'antennes réparties géographiquement pour fournir une meilleure couverture.

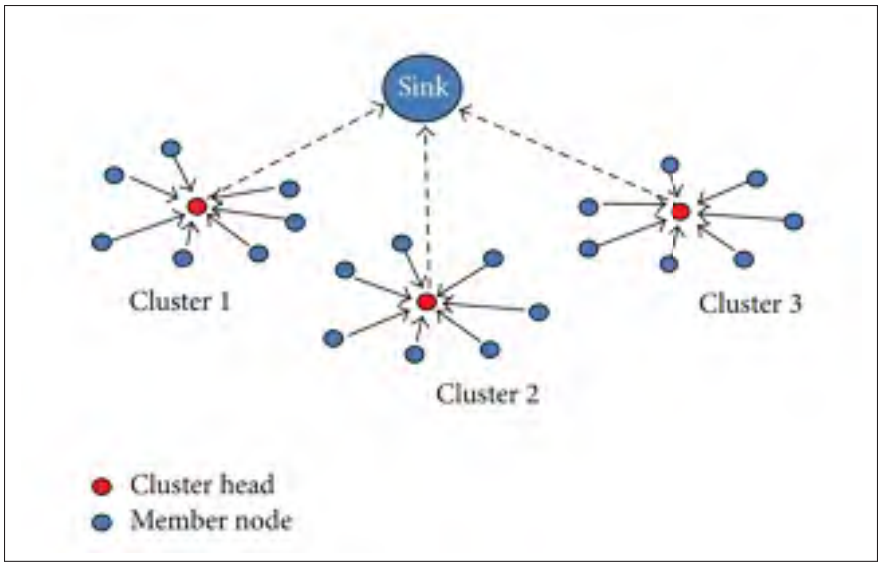

Figure 3.1 Structure des nœuds dans le protocole de routage LEACH

Chaque antenne est reliée à un RRH par l'intermédiaire d'un câble coaxial, et chaque RRH est reliée à une BBU via une fibre optique qui a une perte très faible. Le Fronthaul est la partie entre les RRHs et le BBU pool qui est la partie physique tandis que la partie entre le BBU pool et le réseau central mobile (internet, ressources cloud computing...) est appelée Backhaul qui est la partie virtuelle du réseau.

Le ML-IDS proposé collecte le trafic mobile circulant entre les MSs et le FrontHaul et le traite directement avec le premier modèle qui s'appelle un Perceptron multicouche (MLP). Notre modèle classifie le trafic donné en cinq catégories (classes) ; brouillage constant, brouillage aléatoire, brouillage trompeur, brouillage réactif, brouillage réactif et trafic normal. Dans le cas d'un trafic normal, il sera traité avec un second classificateur, une Machine à Vecteurs de Support Kernelisée (KSVM) pour une deuxième vérification afin de déterminer s'il s'agit vraiment d'un trafic normal ou une attaque de brouillage échappée. La motivation d'ajouter un modèle KSVM après le modèle MLP est de réduire les faux négatifs que le MLP créés. un faux négatif à lieu lorsque le système décide que la situation est normale alors qu'en réalité il y a une attaque.

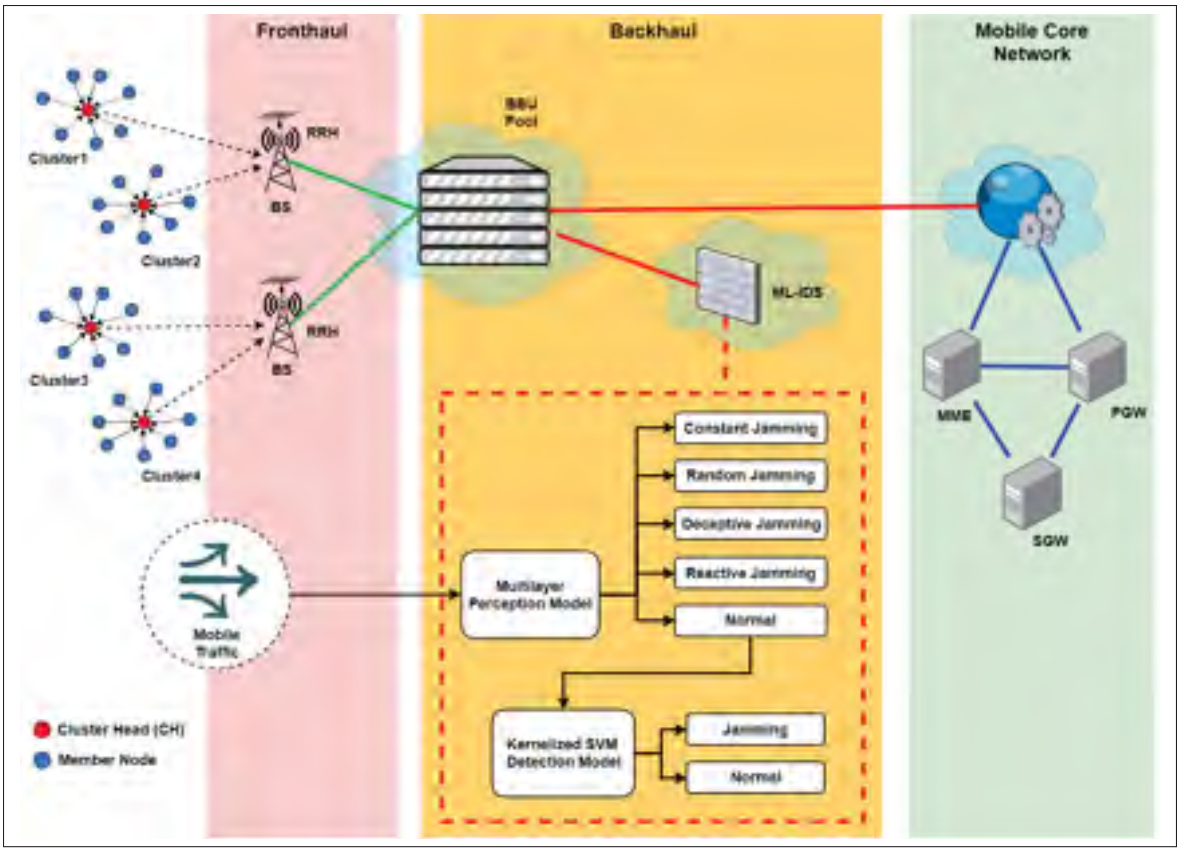

Figure 3.2 Architecture de déploiement du ML-IDS proposé dans les environnements C-RAN

Le ML-IDS est déployé dans le pool de BBU Virtualisé pour plusieurs raisons. Tout d'abord, le pool de BBU virtualisé contient toutes les fonctionnalités du réseau C-RAN telles que l'allocation du spectre, les données confidentielles des utilisateurs, le découpage du réseau, la gestion des services cloud, etc. Ainsi, le pool BBU contrôle l'ensemble du réseau C-RAN. Deuxièmement, l'objectif d'une attaque de brouillage est d'ajouter du bruit dans la zone située entre les regroupements et les stations de base où le pool BBU est la seule partie du C-RAN qui contrôle cette zone. Enfin, le pool BBU est l'infrastructure qui contient suffisamment de ressources pour faire fonctionner un tel moteur de détection sans trop affecter la latence.

Pour obtenir des résultats expérimentaux, un ensemble de données spécialisées pour les réseaux de capteurs sans fil (WSN) a été utilisé pour classer les attaques de brouillage, la WSN-DS contient un trafic réseau normal et malveillant.

3.2 Description de la base de données WSN-DS

Le WSN-DS est un ensemble de données sans fil dédié à la détection d'intrusion (Imen *et al.* (2016)). Il contient exactement 374 661 vecteurs de connexion simples, chacun d'entre eux comprend 23 caractéristiques et est étiqueté comme normal ou attaque. Les types spécifiques d'attaques sont regroupés en différentes catégories, qui sont brouillage constant, brouillage aléatoire, brouillage trompeur, et brouillage réactif, en plus du cas normal (sans attaque).

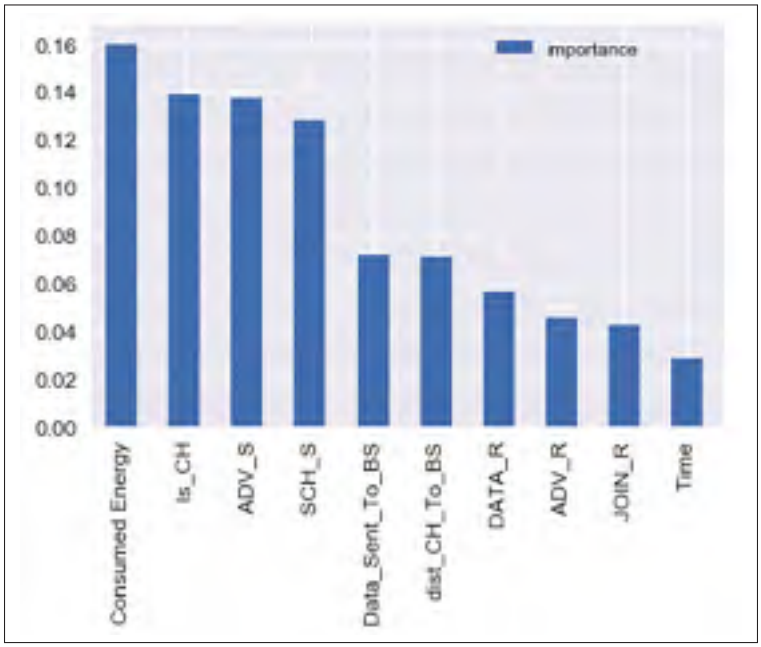

Figure 3.3 Les 10 caractéristiques les plus importantes de WSN-DS

Le WSN-DS contient 23 attributs (caractéristiques) pour aider à déterminer l'état de chaque nœud du réseau C-RAN. Le filtre à faible variance (Low Variance Filter) est un algorithme utile pour la réduction de la dimensionnalité. Les colonnes de données dont les données ont peu varié contiennent peu d'informations. Ainsi, toutes les colonnes de données ayant une variance inférieure à un seuil donné peuvent être supprimées. Notez que la variance dépend de l'étendue de la colonne, et qu'une normalisation est donc nécessaire avant d'appliquer cette technique. Comme nous pouvons le voir dans la figure 3.3, nous avons pu extraire les éléments les plus importants des 23. Les dix éléments suivants ont été sélectionnés :

Energy consumption : la quantité d'énergie consommée lors du cycle précédent.

Dans un premier temps, chaque nœud génère un nombre arbitraire compris entre 0 et 1, puis un seuil est calculé T(n) en utilisant la formule ci-dessous. Si le nombre aléatoire choisi est inférieur à la valeur seuil, le nœud deviendra un CH :

$$T(n) = \begin{cases} \frac{p}{1-p\times(r \bmod p^{-1})} & \forall n \in N \\ 0 & \text{otherwise} \end{cases}$$

où p est la probabilité de CH, N est l'ensemble des nœuds qui n'ont pas été un CH, dans le dernier $1/p$ cycles et r est le cycle actuel.

Is CH : Indicateur permettant de distinguer si le nœud est un CH (valeur 1) ou un nœud normal (valeur 0).

ADV CH send : le nombre de messages publicitaires diffusés par CH envoyés aux nœuds.

ADV SCH send : le nombre de messages de diffusion de grille horaire TDMA envoyés aux nœuds.

Data sent to BS : la quantité de paquets de données transmis au RRH.

Distance CH to BS : la distance entre le CH et le RRH.

Data received : le nombre de paquets reçus de CH.

ADV CH receives : le nombre de messages publicitaires de CH reçus des CH.

Join REQ receive : le nombre de messages de demande de jointure reçus par les CH des nœuds.

Time : le temps de simulation actuel du nœud.

3.3 Algorithmes implémentés

Deux classificateurs ont été implémentés dans le pool BBU pour détecter les quatre types d'attaques de brouillage pour un contrôle à plusieurs niveaux (Garg, S. (2020)). Par conséquent, si une attaque est ratée, le second pourra la détecter.

Le premier classificateur que nous avons utilisé est un classificateur d'apprentissage profond (Singh *et al.* (2020)) qui s'appelle le MLP. La raison pour laquelle on a utilisé ce classificateur est que le nombre des vecteurs d'entrée que nous avons est très grand (dans notre cas c'est 374661 vecteurs), donc la baisse du gradient stochastique est souvent le meilleur choix (surtout pour la classification) en termes de vitesse et de capacité. Et le MLP a été choisi parmi tous les algorithmes d'apprentissage profond (CNN, RNN, LSTM, DBM, DBM, DBN ...) pour sa flexibilité qui lui a permet d'être appliquée à d'autres types de données.

Le deuxième classificateur implémenté est la KSVM qui est un classificateur discriminant. La raison pour laquelle on a utilisé ce dernier c'est qu'il est le meilleur dans la classification binaire.

3.3.1 MLP

Le perceptron multicouche est constitué d'un système de neurones simples interconnectés, ou noeuds, comme illustré dans la figure 3.4 Tirée de WGardnera *et al.* (1998), qui est un modèle représentant une cartographie non-linéaire entre un vecteur d'entrée et un vecteur de sortie. Les nœuds sont reliés par des poids et des signaux de sortie qui sont une fonction de la somme des entrées du nœud modifiées par une simple fonction de transfert non-linéaire, ou d'activation. C'est la superposition de nombreuses fonctions non-linéaire qui permettent au perceptron multicouche de se rapprocher de fonctions extrêmement non-linéaires.

Si la fonction de transfert était linéaire, le perceptron multicouche ne pourrait modéliser que des fonctions linéaires. En raison de sa dérivée facile à calculer, une fonction de transfert couramment utilisée est la fonction logistique, comme le montre la Fig. 3.4 la sortie d'un nœud

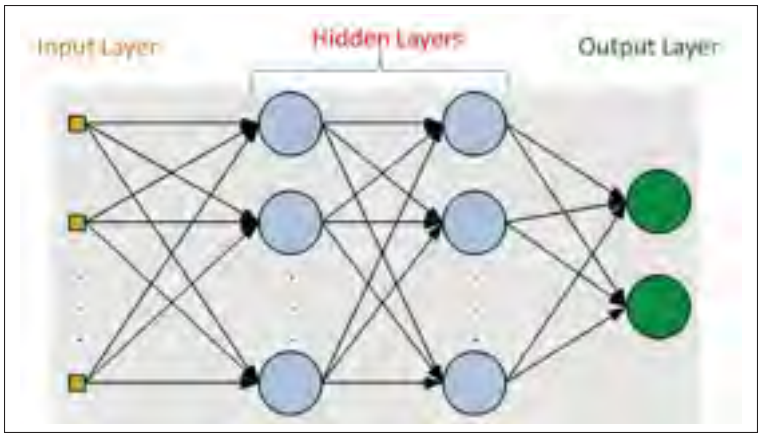

Figure 3.4 Un perceptron multicouche avec des couches cachées

est mise à l'échelle par le poids de connexion et alimentée vers l'avant pour être une entrée vers les nœuds de la couche suivante du réseau. Cela implique une direction du traitement de l'information, d'où le perceptron multicouche connu sous le nom de réseau neuronal de feed-forward. L'architecture d'un perceptron multicouche est variable, mais se compose généralement de plusieurs couches de neurones. La couche d'entrée ne joue aucun rôle de calcul, mais elle sert à transmettre le vecteur d'entrée au réseau. Les vecteurs d'entrée et de sortie se réfèrent aux entrées et sorties du perceptron multicouche et peuvent être représentés comme des vecteurs simples, comme le montre la figure 3.5 un perceptron multicouche peut avoir une ou plusieurs couches cachées.

3.3.2 KSVM

La KSVM est un classificateur discriminant, sa puissance réside dans le fait qu'il utilise une fonction noyau qui mappe les données dans un espace différent où un hyperplan linéaire peut être utilisé pour classifier les attaques comme on peut le voir dans la Figure 3.6 Tirée de Rojo-lvarez *et al.* (2018). C'est ce qu'on appelle le *kernel trick* où la fonction kernel transforme les données dans l'espace dimensionnel supérieur de la caractéristique de sorte qu'une séparation linéaire est possible.

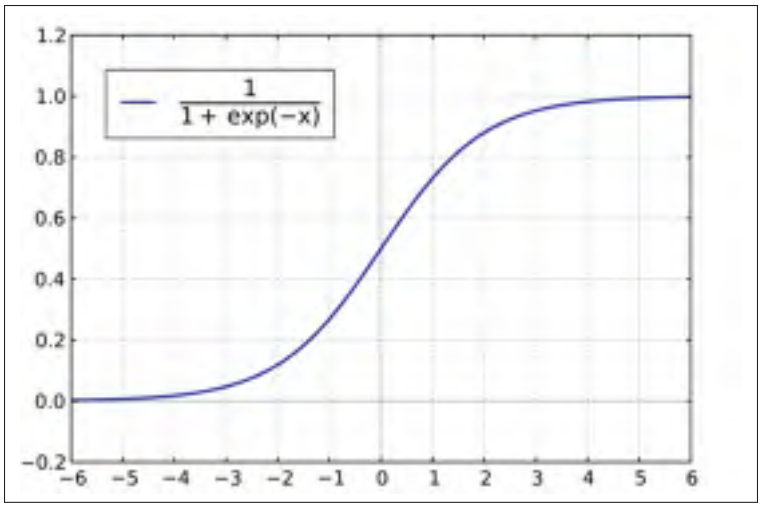

Figure 3.5 La fonction logistique

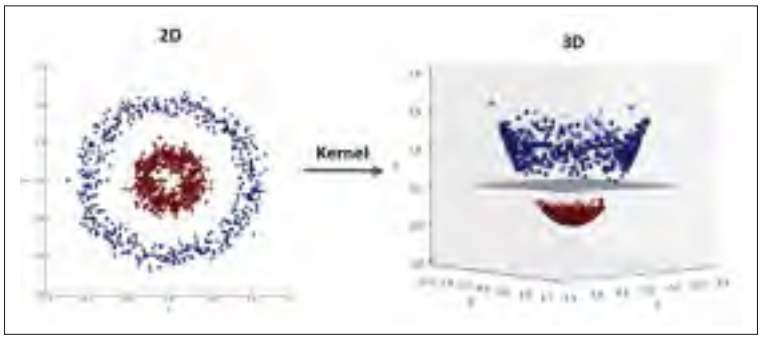

Figure 3.6 Classificateur non-linéaire utilisant l'astuce du Kernel

La limite de décision ou hyperplan séparant les classes a des coefficients de pondération donnés par le vecteur W, que nous devons estimer. Le classificateur SVM essaie de maximiser la distance entre ce vecteur W et les points les plus proches (vecteurs de support) pour qu'il devienne notre contrainte. Cela équivaut à minimiser l'équation suivante :

$$W(\alpha) = -\sum_{i=1}^{l} \alpha_i + \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j x_i x_j$$

Sachant que :

$$\sum_{i=1}^{l} y_i \alpha_i = 0$$

l est le nombre de points de données dans nos données d'entraînement, *y* indique les sorties des points de données, *x* est le vecteur caractéristique dans chaque exemple d'entraînement, et α est la constante lagrangienne.

3.4 Matrice d'évaluation de performance

Pour évaluer l'approche proposée, nous utilisons la matrice d'évaluation de performance et évaluons *l'éxactitude, la précision, le rappel, et la F-mesure*. Nous notons que TP (true positive) est le nombre d'exemples d'attaques correctement classées, TN (true negative) est le nombre d'exemples normaux (no attack) correctement classés comme normaux, FP (false positive) est le nombre d'exemples normaux mal classés comme attaques, FN (false negative) est le nombre d'exemples d'attaques mal classées comme normales.

3.4.1 Éxactitude

C'est la mesure de performance la plus intuitive et il s'agit simplement d'un rapport entre les TPs et le nombre total d'observations. On peut penser que, si nous avons une grande précision, notre modèle est le meilleur. En effet, l'exactitude est une grande mesure, mais seulement lorsque nous avons des ensembles de données symétriques où les valeurs des FPs et des FNs sont presque identiques. Par conséquent, vous devez tenir compte d'autres paramètres pour évaluer la performance de votre modèle :

$$Exactitude = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

3.4.2 Précision

C'est le rapport entre le nombre d'observations positives correctement prévues TPs et le nombre total d'observations positives prévues :

$$Prcision = \frac{TP}{TP+FP} \tag{3.2}$$

3.4.3 Rappel

C'est le rapport entre les observations positives correctement prédites TPs et toutes les observations de la classe réelle :

$$Rappel = \frac{TP}{TP+FN} \tag{3.3}$$

3.4.4 F-mesure

C'est la moyenne pondérée de la Précision et du Rappel. Par conséquent, ce score tient compte à la fois des FPs et des FNs. Intuitivement, il n'est pas aussi facile à comprendre que la précision, mais le F-mesure est généralement plus utile que la précision, surtout si vous avez une distribution inégale des classes :

$$F-mesure = F1score = \frac{TP}{TP+FP+FN} \tag{3.4}$$

3.5 Conclusion

Dans ce chapitre, nous avons bien présenté notre solution proposée, ainsi que l'architecture de déploiement dans le réseau C-RAN. Nous avons ainsi détaillé les algorithmes que nous allons implementer par la suite.

CHAPITRE 4

SIMULATIONS ET RÉSULTATS

4.1 Introduction

Dans ce chapitre, nous allons évaluer la performance des algorithmes définis dans le chapitre précédent afin de détecter et classifier quatre types d'attaques de brouillage : brouillage constant, brouillage aléatoire, brouillage trompeur, et brouillage réactif. Nous allons finir par une comparaison des faux négatifs.

4.2 Environnement de la simulation

Toutes les tâches sont effectuées à l'aide du langage de programmation Python et la bibliothèque Scikit-learn. Nous avons exécuté notre code source en utilisant le programme Jupyter Notebook qui est une application web open source qui permet de combiner facilement Markdown texte et code source Python exécutable sur un canevas appelé un carnet de notes. Les expériences ont été menées sur un processeur Intel(R) Xeon(R) CPU E3-1225 v5 @ 3.30 GHz 3.31 GHz, 16.00 Go de RAM avec Windows 10 Enterprise 2016 LTSB 64 bits Système d'exploitation et processeur x64.

4.3 Résultats

Dans cette partie, on va présenter les résultats des classificateurs implémentés ainsi que la précision de chaque algorithme.

4.3.1 Séparation de la base de données WSN-DS

L'objectif fondamental de l'apprentissage par machine est de former des modèles généraux au-delà des instances de données utilisées. Nous souhaitons évaluer le modèle afin d'estimer la qualité de la généralisation de son modèle pour les données sur lesquelles le modèle n'a pas

été formé. Cependant, étant donné que les instances futures ont des valeurs cibles inconnues et que nous ne pouvons pas vérifier l'exactitude de nos prévisions pour les instances futures à présent, nous devons utiliser certaines des données pour lesquelles nous connaissons déjà la réponse comme proxy pour les données futures. Évaluer le modèle avec les mêmes données que celles utilisées pour la formation n'est pas utile, car il récompense les modèles qui peuvent «mémoriser» les données de la formation, au lieu de les généraliser.

Une stratégie courante consiste à prendre toutes les données étiquetées disponibles et à les diviser en sous-ensembles d'entraînement et d'évaluation, généralement avec un ratio de 70% pour l'entraînement et 30% pour l'évaluation. Le système ML utilise les données d'apprentissage pour entraîner les modèles à observer les modèles et utilise les données d'évaluation pour évaluer la qualité prédictive du modèle formé. Le système ML évalue la performance prédictive en comparant les prédictions de l'ensemble de données d'évaluation avec des valeurs vraies en utilisant diverses mesures.

La WSN-DS a été divisé en 70% de données d'entraînement et 30% de données d'évaluation. Le tableau I montre la séparation des données.

Tableau 4.1 Nombre d'enregistrements utilisés dans les ensembles de données d'entrainement et d'évaluation.

Classes	**Nombre d'enregistrements utilisés dans la WSN**	
	données d'entraînement (70%)	**données d'évaluation (30%)**
Normal	238103	101963
Brouillage constant	10233	4363
Brouillage aléatoire	6960	3089
Brouillage trompeur	4650	1988
Brouillage réactif	2316	996
Somme	**262262**	**112399**

4.3.2 Précision des modèles implémentés

Deux classificateurs ont été mis en œuvre, notamment le MLP et la KSVM. Le tableau II montre la précision obtenue pour chaque classificateur mis en œuvre.

Tableau 4.2 Précision des modèles de classification d'attaques.

Méthodes utilisées	**Précision**
MLP (Premier niveau)	81.73 %
MLP + KSVM (Deuxième niveau)	94.51 %

4.3.3 Comparaison avec d'autres modèles

Pour évaluer l'approche proposée, nous avons comparé notre modèle en plusieurs étapes avec un autre travail qui utilisait uniquement le MLP (Imen *et al.* (2016)) appliqué au même ensemble de données pour confirmer que notre approche est la plus appropriée. Le tableau 4.3 montre que notre modèle peut fournir une meilleure précision de classification d'attaques que le modèle MLP tout seul.

Tableau 4.3 Précision globale et pour chaque classe.

Modèle	**Précision de la classification d'attaques**						**Précision globale**
	Rand.	**Const.**	**React.**	**Decept.**	**jamm.**	**Norm.**	
MLP (Imen *et al.* (2016))	92.8%	75.6%	99.4%	92.2%	-	99.8%	**91.9%**
MLP + KSVM	95.3%	82.9%	99.6%	94.7%	98.2%	100%	**94.5%**

4.3.4 Faux négatifs et faux positifs

Pour évaluer l'approche proposée, le tableau 4.4 montre le nombre de faux positifs et de faux négatifs à la sortie de chaque classificateur.

Tableau 4.4 Faux négatifs et faux positifs.

Classificateur	FN/FP	Percentage of FN/FP	Numbers of FN/FP
MLP	**Faux Négatifs**	7,98 %	8969
	Faux Positifs	11,04 %	12363
MLP + KSVM	**Faux Négatifs**	7,84 %	8812
	Faux Positifs	10,95 %	12307

Sachant que les faux négatifs sont les plus pires erreurs qu'un modèle peut générer, nous pouvons voir dans le tableau précédent que le deuxième classificateur KSVM a réussi à sauver notre architecture C-RAN de 157 FNs qui ont été manquées par le MLP. Par conséquent, une plus grande précision de classification et une meilleure sécurité dans les architectures C-RAN sont obtenues.

4.3.5 Courbe ROC

La courbe des caractéristiques de fonctionnement du récepteur (ROC) est utilisée pour visualiser les performances des classificateurs. Elle nous donne le compromis entre le taux de vrai positif (TPR) et le taux de faux positif (FPR) à différents seuils de classification.

$$TPR = \frac{TP}{TP+FN}$$

$$FPR = \frac{FP}{TN+FP}$$

Comme le montrent les équations ci-dessus, le TPR est la proportion d'observations dont on prévoit correctement qu'elles seront positives. Cependant, le FPR est la proportion d'observations dont on prédit à tort qu'elles seront positives. La figure 4.1 montre la courbe ROC pour chaque modèle.

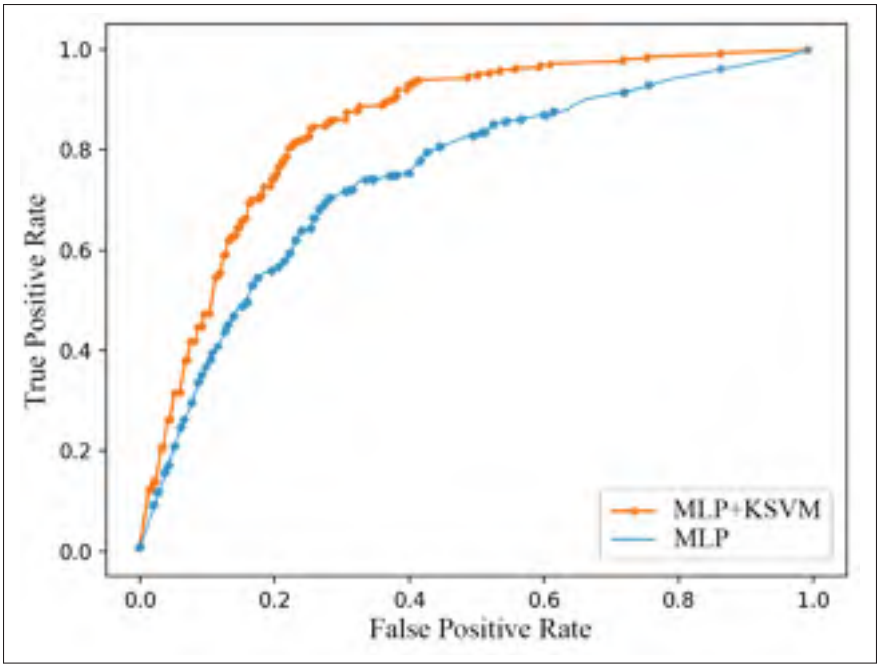

Figure 4.1 La courbe des caractéristiques de fonctionnement du récepteur (ROC)

4.3.6 Matrice d'évaluation de performance

Le tableau 4.5 montre la précision, le recall, le F1-score et le support pour les cinq classes : brouillage constant, brouillage aléatoire, brouillage trompeur, et brouillage réactif et le cas normal (sans attaques).

Tableau 4.5 Résultats des paramètres de la matrice d'évaluation de performance

Types d'attaques	**Matrice d'évaluation de performance**			
	Precision	**Recall**	**F1-score**	**Support**
Brouillage constant	0.99	0.96	0.97	3058
Brouillage aléatoire	0.98	0.98	0.98	4276
Brouillage trompeur	0.99	1.00	0.99	986
Brouillage réactif	0.97	1.00	0.98	1928
Normal	1.00	1.00	1.00	101963
Avg/total	***0.98***	***0.98***	***0.98***	***112211***

4.4 Discussions

Sur la base des résultats obtenus ci-dessus, nous pouvons conclure les avantages de la ML-IDS à plusieurs niveaux que nous avons proposés :

En termes de précision de classification : Nous avons proposé un ML-IDS à plusieurs niveaux basé sur un apprentissage supervisé et profond pour la détection et la classification d'attaques de brouillage avec une haute précision.

En termes d'implémentation : Nous avons concevé, implémenté et déployé ce nouveau ML-IDS dans le pool du BBU virtualisé au sein de l'architecture du C-RAN de manière à ne pas trop affecter la latence.

En termes de la sécurité : Nous avons amélioré la sécurité des réseaux C-RAN en détectant quatre types d'attaques de brouillage, qui sont les plus populaires et les plus graves dans les réseaux sans fil.

4.5 Conclusion

À partir des résultats de simulations, nous déduisons que le classificateur MLP à deux couches cachées génére plusieurs faux négatifs et positives qui peuvent conduire à une mauvaise classification d'attaques. Ensuite, à l'aide du KSVM, nous avons réussi à sauver notre architecture C-RAN en détectant 157 FNs et 56 FPs qui ont été manquées au début par le MLP.

CONCLUSION ET RECOMMANDATIONS

Dans ce mémoire, nous avons abordé l'un des principaux défis auquel sont confrontés les réseaux C-RAN qui seront implémentés en 2020 par la 5G. Ce défi consiste à détecter et classer quatre types d'attaques de brouillage qui représentent les menaces de sécurité les plus sérieuses dans le domaine des communications sans fil.

Nous avons d'abord passé en revue certains des principes de base, des avantages et des défis actuels de la technologie C-RAN qui prennent en charge la virtualisation des stations de base dans les futurs réseaux cellulaires. Ensuite, nous avons présenté les systèmes de détection d'intrusion (IDS) et les types d'IDS qui existent, ainsi que les différentes méthodes de détection afin de détecter les quatre types d'attaques de brouillage qui affrontent les réseaux C-RAN.

Ensuite, nous avons déployé notre ML-IDS à plusieurs niveaux au sein de l'architecture C-RAN, puis on a réussi à implémenter deux algorithmes d'apprentissage supervisé et profond afin d'approuver notre solution. Les résultats ont démontré que le deuxième niveau de détection, qui est le classificateur KSVM, a réussi à sauver notre architecture C-RAN de plus de 200 attaques (157 FNs et 56 FPs) qui ont été manquées par le premier niveau de détection, le classificateur MLP. Les résultats montrent que notre déploiement assure : 1) une sécurité accrue, 2) un déploiement solide au cœur de l'architecture C-RAN et 3) une grande précision dans la détection d'attaques.

Dans une perspective future, on compte créer notre propre base de données sans fil dédié à la détection d'intrusion afin d'inclure d'autres types d'attaques de brouillage telles que le brouillage intelligent basé sur le bruit de tir. En outre, plusieurs attaques qui ciblent les architectures C-RAN comme les attaques d'écoute, les attaques d'émulation d'utilisateur primaire et les attaques d'usurpation d'identité seront incluses.

BIBLIOGRAPHIE

A. Abeshu, and N. Chilamkurti. (2018). *Deep learning : the frontier for distributed attack detection in fog-to-things computing*. IEEE Communications Magazine 56.2 : 169-175.

A. Ghorbani, W. Lu, and T. Mahbod. (2009). *Network intrusion detection and prevention : concepts and techniques*. Vol. 47. Springer Science Business Media.

A. Honig, H. Andrew, E. Eleazar,and S. Salvatore. (2002). *Adaptive model generation : An architecture for the deployment of data mining based intrusion detection systems*. Data Mining for Security Applications. Boston, Kluwer Academic Publishers.

A. Iman , A. Bassam, and A. Mousa . (2016). *Wsn-ds : A dataset for intrusion detection systems in wireless sensor networks*. Journal of Sensors, 16 pages.

A. Jain, B.Verma, and J. L. Rana. (2014). *Anomaly intrusion detection techniques : A brief review*. Journal of Scientific Engineering Research, Vol. 5, Iss. 7, pp.1372-1383.

A. Midzic, Z. Avdagic, and S. Omanovic. (2016). *Intrusion detection system modeling based on neural networks and fuzzy logic*. IEEE 20th Jubilee International Conference on Intelligent Engineering Systems (INES).

A. Mpitziopoulos, D. Gavalas, C. Konstantopoulos, and G. Pantziou. (2009). *A survey on jamming attacks and countermeasures in wsns*. IEEE Communications Surveys anf Tutorials, vol. 11, no. 4, pp. 42-56, Fourth Quarter.

A. Singh, G. S. Aujla, S. Garg, G. Kaddoum and G. Singh,title = Deep Learning-based SDN Model for Internet of Things : An Incremental Tensor Train Approach, p. . i. y. . .

B. Manu. (2016). *A survey on secure network : Intrusion detection prevention approaches*. American Journal of Information Systems, vol. 4, no. 3 (2016) : 69-88.

C. M. Moreira, G. Kaddoum and E. Bou-Harb. (2018). *Cross-layer authentication protocol design for ultra-dense 5g hetnets*. 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, 2018, pp. 1-7, doi : 10.1109/ICC.2018.8422404.

C. Miranda, G. Kaddoum, E. Bou-Harb, S. Garg and K. Kaur. (2020). *A collaborative security framework for software-defined wireless sensor networks*. IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2602-2615, 2020, doi : 10.1109/TIFS.2020.2973875.

C. Puñal, I. Aktaş, C. Schnelke, G. Abidin, K. Wehrle, and J. Gross. (2014). *Machine learning-based jamming detection for ieee 802.11 : Design and experimental evaluation*. Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014, Sydney, NSW, pp. 1-10.

C. Varun, B. Arindam, and K. Vipin. (2009). *Anomaly detection : A survey*. ACM Computing Surveys, Vol. 41, No. 3, Article 15.

D. Karagiannis, and A. Argyriou. (2018). *Jamming attack detection in a pair of rf communicating vehicles using unsupervised machine learning*. Vehicular Communications, Volume 13, Pages 56-63.

D. Khan, Z. Ali, and P. Herrmann. (2017). *A trust based distributed intrusion detection mechanism for internet of things*. Advanced Information Networking and Applications (AINA), IEEE 31st International Conference on. IEEE.

E. Dickerson, D. Julie. (2000). *Fuzzy network profiling for intrusion detection*. International Conference of the North American Fuzzy Information Processing Society (NAFIPS), Atlanta, GA, pp. 301, 306.

E. Eskin, A. Andrew, P. Michael, P. Leonid, and S. Sal. (2002). *A geometric framework for unsupervised anomaly detection : Detecting intrusions in unlabeled data*. Data Mining for Security Applications, Boston : Kluwer Academic Publishers.

E. Hodo. (2016). *Threat analysis of iot networks using artificial neural network intrusion detection system*. Networks, Computers and Communications (ISNCC), International Symposium on. IEEE.

F. Jameel, S. Wyne, G. Kaddoum and T. Q. Duong. (2019). *A comprehensive survey on cooperative relaying and jamming strategies for physical layer security*. IEEE Communications Surveys Tutorials, vol. 21, no. 3, pp. 2734-2771, thirdquarter 2019, doi : 10.1109/COMST.2018.2865607.

F. Tian, P. Zhang, and Z. Yan. (2017). *A survey on c-ran security*. IEEE Access, vol. 5, pp. 13372-13386.

G. Han, L. Xiao, and H. Poor. (2017). *Two-dimensional anti-jamming communication based on deep reinforcement learning.* ieee international conference on acoustics, speech and signal processing (icassp), pp. 2087–2091.

I. Parvez, A. Rahmati, I. Güvenç, and H. Dai. (2018). *A survey on low latency towards 5g : Ran, core network and caching solutions*. IEEE Communications surveys and Tutoriels, Vol. 20, No. 4, Fourth Quartier.

J. Huang, R. Duan, C. Cui, J. Jiang, and L. Liam. (2014). *Recent progress on c-ran centralization and cloudification*. IEEE Access, vol. 2, pp. 1030-1039.

J. McHugh,. (2001). *Intrusion and intrusion detection*. International Journal of Information Security 1 (2001), no. 1, 14–35.

J. Yang, Y. Qiao, X. Zhang, H. He, F. Liu, and G. Cheng. (2015). *Characterizing user behavior in mobile internet*. IEEE Transactions on Emerging Topics in Computing, vol. 3, no. 1, pp. 95-106.

K. Bajaj, and A. Arora. (2013). *Improving the intrusion detection using discriminative machine learning approach and improve the time complexity by data mining feature selection methods.* International journal of computer applications, 76.

K. Boulos, M. El Helou, and S. Lahoud. (2015). *Rrh clustering in cloud radio access networks*. 2015 International Conference on Applied Research in Computer Science and Engineering (ICAR), Beirut, pp. 1-6.

L. Han, M. Zhou, W. Jia, Z. Dalil, and X. Xu. (2019). *Intrusion detection model of wireless sensor networks based on game theory and an autoregressive model*. Information sciences, 476, 491–504.

L. Kong, J. He, G. Kaddoum, S. Vuppala and L. Wang. (2016). *Secrecy analysis of a mimo full-duplex active eavesdropper with channel estimation errors*. 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), Montreal, QC, 2016, pp. 1-5, doi : 10.1109/VTC-Fall.2016.7881216.

M. Atallah and G. Kaddoum. (2019). *Secrecy analysis in wireless network with passive eavesdroppers by using partial cooperation*. IEEE Transactions on Vehicular Technology, vol. 68, no. 7, pp. 7225-7230, July 2019, doi : 10.1109/TVT.2019.2913934.

M. Atallah, G. Kaddoum and L. Kong. (2015). *A survey on cooperative jamming applied to physical layer security*. IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB), Montreal, QC, 2015, pp. 1-5, doi : 10.1109/ICUWB.2015.7324413.

M. Hachimi, G. Kaddoum, G. Gagnon, and P. Illy. (2020). *Multi-stage jamming attacks detection using deep learning combined with kernelized support vector machine in 5g cloud radio access networks*. IEEE International Symposium on Networks, Computers and Communications (ISNCC'20), arXiv :2004.06077.

M. Hadzialic, B. Dosenovic, M. Dzaferagic, and J. Musovic. (2013). *Cloud-ran : Innovative radio access network architecture*. Proceedings ELMAR-2013, Zadar, 2013, pp. 115-120.

M. Tavallaee. (2009). *A detailed analysis of the kdd cup 99 data set*. Computational Intelligence for Security and Defense Applications. CISDA 2009. IEEE Symposium on. IEEE.

N. Neshenko, E. Bou-Harb, J. Crichigno, G. Kaddoum and N. Ghani. (2019). *Demystifying iot security : An exhaustive survey on iot vulnerabilities and a first empirical look on internet-scale iot exploitations*. IEEE Communications Surveys Tutorials, vol. 21, no. 3, pp. 2702-2733, thirdquarter 2019, doi : 10.1109/COMST.2019.2910750.

P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, . (2009). *Anomalybased network intrusion detection : Techniques, systems and challenges.* computers security, 28(1-2), 18–28.

P. Garcia-Teodoro, J. Diaz-Verdejo, M. Gabriel, and V. Enrique . (2009). *Anomaly-based network intrusion detection : Techniques, systems and challenges*. computers security, vol.28, no. 1, pp. 18, 28.

P. Illy. (2018a). *Sécurité dans les réseaux iot 5g : Détection et prévention d'intrusion.* 2018 l'École de Technologie Supérieure.

P. Illy. (2018b). *Les systèmes de détection d'intrusion (ids). rapport de recherche pour le cours ift6271-sécurité informatique*. 2018 l'Université de Montréal.

P. Illy, G. Kaddoum, C.M. Moreira, K. Kaur, and S. Garg. (2019). *Securing fog-to-things environment using intrusion detection system based on ensemble learning*. 2019 IEEE Wireless Communications and Networking Conference (WCNC).

R. Das, A. Gadre, S. Zhang, S. Kumar, and J. M. Moura. (2018). *A deep learning approach to iot authentication.* ieee international conference on communications (icc), pp. 1–6.

S. Buzzi, T. E. Klein, H. Poor, C. Yang, and A. Zappone. (2016). *A survey of energy-efficient techniques for 5g networks and challenges ahead*. IEEE Journal on Selected Areas in Communications, vol. 34, no. 4, pp. 697-709.

S. Danish, A. Nasir, H. K. Qureshi, A. B. Ashfaq, S. Mumtaz, and J. Rodriguez. (2018). *Network intrusion detection system for jamming attack in lorawan join procedure*. IEEE International Conference on Communications (ICC), Kansas City, MO.

S. Garg, K. Kaur, G. Kaddoum, F. Gagnon, N. Kumar, and Z. Han. (2019). *Sec-iov : A multi-stage anomaly detection scheme for internet of vehicles*. ACM MobiHoc Workshop on Pervasive Systems in the IoT Era (PERSIST-IoT '19). Association for Computing Machinery, New York, NY, USA, 37–42. Doi : 10.1145/3331052.3332476.

S. Garg, K. Kaur, G. Kaddoum, S. H. Ahmed and D. N. K. Jayakody. (2019). *Sdn-based secure and privacy-preserving scheme for vehicular networks : A 5g perspective*. IEEE Transactions on Vehicular Technology, vol. 68, no. 9, pp. 8421-8434, Sept. 2019, doi : 10.1109/TVT.2019.2917776.

S. Garg, K. Kaur, N. Kumar, G. Kaddoum, A. Y. Zomaya and R. Ranjan. (2019). *A hybrid deep learning-based model for anomaly detection in cloud datacenter networks*. IEEE Transactions on Network and Service Management, vol. 16, no. 3, pp. 924-935, Sept. 2019, doi : 10.1109/TNSM.2019.2927886.

S. Garg, K. Kaur, S. Batra, G. Kaddoum, N. Kumar, and A. Boukerche. (2020). *A multi-stage anomaly detection scheme for augmenting the security in iot-enabled applications*. Future Gener. Comput. Syst., vol. 104, pp. 105–118, Mar. 2020, doi : 10.1016/j.future.2019.09.038.

S. Kumar. (2014). *Technique for security of multimedia using neural network*. International Journal of Research in Engineering Technology and Management, vol. 2, issue 5, pp.1-7.

S. Mavoungou, G. Kaddoum, M. Taha and G. Matar. (2016). *Survey on threats and attacks on mobile networks*. IEEE Access, vol. 4, pp. 4543-4572, 2016, doi : 10.1109/ACCESS.2016.2601009.

T. Heer, O. Garcia-Morchon, and K. Wehrle. (2011). *Security challenges in the ip-based internet of things*. Wireless personal communications, 61(3), 527–542.

T. Le, T. Park, D. Cho, and H. Kim. (2018). *An effective classification for dos attacks in wireless sensor networks*. Tenth International Conference on Ubiquitous and Future Networks (ICUFN), Prague, pp. 689-692.

W. Hong, and H. Jiang. (2017). *Multibeam antenna technologies for 5g wireless communications*. Ieee transactions on antennas and propagation, 65(12), 6231–6249.

Y. Shi, Y. E. Sagduyu, and J. H. Li. (2018). *Adversarial deep learning for cognitive radio security : Jamming attack and defense strategies*. IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, MO, pp. 1-6.