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 :

BOUAZZA Abdelhammid

CHAABI Aissa

Sur le thème

Détection des attaques de routage dans l'Internet des Objets

Soutenu publiquement à Tiaret devant le jury composé de :

Mr. KHARROUBI Sahraoui	M.C.B	Université IBN-KHALDOUN Tiaret	Président
Mr. ALEM Abdelkader	M.A.A	Université IBN-KHALDOUN Tiaret	Encadreur
Mr DAOUD Mohamed Amine	M.A.A	Université IBN-KHALDOUN Tiaret	Examinateur

2019 - 2020

Remerciements

Nous tenons tout d'abord à remercier DIEU le tout puissant et

Le miséricordieux, qui nous a donné la force

et la patience d'accomplir ce modeste travail

Un très grand merci à :

Nos parents qui nous ont suivis pendant nos études.

En second lieu nous tenons à remercier notre encadreur

Mr ***ALEM Abdelkader*** *pour son aide, pour son encouragement,*

et pour ses précieux conseils durant la réalisation de ce travail.

Nos vifs remerciements vont également aux membres du jury

Mr ***KHARROUBI Sahraoui*** *et Mr* ***DAOUD Mohamed Amine***

qui ont pris de leur temps pour juger ce modeste travail,

qu'ils trouvent ici l'expression de notre gratitude et tout notre respect.

Nous adressons aussi nos remerciements à tous les professeurs qui

nous ont enseignés durant ce cursus universitaire.

En fin, nous remercions Nos collègues de la promotion ***2018-2020****.*

Tout en leurs souhaitons un avenir plein de réussite.

Dédicaces

Je dédie ce modeste travail à :

- *À mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leurs prières tout au long de mes études*
- *A mes sœurs et frères et je les souhaite beaucoup de joie et de bonheur dans la vie.*
- *A tout ma famille.*
- *A tous mes amis et mes camarades.*
- *A tous ceux qui me sont chers.*

- *Je n'oublierai pas de dédier et de remercie mon Directeur Hadj HAMI djamel el dine et mes collègues de travail pour leur compréhension et leur soutien.*

BOUAZZA Abdelhammid

Dédicaces

Je dédie cet humble travail à :

- *Mes chers parents qui m'ont toujours encouragé et soutenu durant tout mon cursus universitaire.*
- *Ma petite famille, ma femme que je ne pourrai jamais remercier assez pour son aide et son soutien ; mes enfants NABIL. SARAH WAFAA ET IMENE.*
- *Mes frères et sœurs.*
- *Tous mes amis.*

CHAABI Aissa

Résumé

L'**internet des Objets** ou IdO (en anglais Internet of Things ou IoT) n'est plus une fantaisie scientifique. La progression technologique permet maintenant d'apercevoir la connexion des objets du quotidien à l'Internet, des solutions ouvertes et interopérables doivent cependant être utilisées pour assurer une communication optimum entre ces objets, le protocole de routage est un élément clé de cet objectif. Dans ce contexte, nous avons étudié et présenté le protocole RPL (Routing Protocol for Low Power Lossy Network) qui est l'un des principaux protocoles de routage. La nature des objets et les contraintes qui posent, sont convaincantes pour le choix du modèle de détection des attaques. Compte tenu de la nature des réseaux RPL, il est obligatoire d'identifier et d'analyser les attaques auxquelles ce protocole est confronté. Plusieurs travaux de recherche ont mis l'accent sur la détection des attaques dans les protocoles de routage. Dans notre PFE, nous proposons un mécanisme d'IDS (Intrusion detection system), Que nous avons appelé : **Minerva-IDS** pour la détection des attaques visant le routage dans réseau IdO et qui se base sur la détection des intrusions par comportement grâce à l'apprentissage automatique, puis nous avons implémenté et simulé ce protocole à l'aide du simulateur Contiki Cooja plusieurs scénarios du réseau. Ensuite, nous avons construit notre ensemble des données en utilisant des paramètres importants pour détecter les attaques de routage dans réseau IdO, qui est nécessaire pour créer notre modèle IDS (hybride et hiérarchique). Parmi les travaux consultés, plusieurs sont basés des classificateurs du même niveau et de façon isolée. Pour cela Nous avons proposé un modèle de 03 niveaux combine entre différentes techniques de classifications (Binaire et Multi-classes) et le filtrage afin de réduire la taille de l'ensemble de données, et de minimiser au maximum de fausses alertes. Les résultats obtenus sont très satisfaisants car on réussit à obtenir un *taux minimum* des fausses alertes (faux positive) et un *taux maximum* de détection d'attaques, avec une réduction du temps d'apprentissage et de prédiction.

Mots clés : Internet Des Objets, Internet Of Things, Système de Détection d'Intrusion, DODAG, RPL, sécurité, routage, attaque …

Abstract

The Internet of Things is no longer a scientific fantasy. Ongoing technological progress enables the connection of everyday objects to the Internet. Open and interoperable solutions must however be used to ensure optimum communication between these objects, and the routing protocol is a key element of this objective.

The RPL(Routing Protocol for Low Power Lossy Network) is one of the main routing protocols. The nature of the objects and the constraints that they pose are convincing for the choice of the attack detection model. Given the nature of RPL networks, it is compulsory to identify and analyze the attacks that this protocol faces.

In this work, we propose a mechanism (Minerva-IDS), for the detection of attacks targeting the routing in IoT network, based on the detection of intrusions by behavior thanks to machine learning.

We determined, using the Contiki Cooja simulator several network scenarios. Then, we built our dataset using important parameters to detect routing attacks in IoT network, which is necessary to create our IDS (Intrusion Detection Systems) hybrid and hierarchical.

We have proposed a 03-level model that combines different classification techniques (Binary and Multi-class) and filtration in order to reduce the size of the dataset, and minimize a maximum of false alarms.

The results obtained were very satisfactory because they reduced the percentage of false alarms (false positive) and maximized the detection rate, and with a reduction of the learning and prediction time.

Keywords : Internet Of Things, ROUTING , Intrusion Detection Systems , DODAG, RPL. SECURITY, ATTACK …

الملخص

لم يعد إنترنت الأشياء خيالًا علميًا. فالتقدم التكنولوجي يجعل الأشياء تتصل فيما بينها يوميا عبر الإنترنت ، ومع ذلك يجب استخدام الحلول المفتوحة والقابلة للتشغيل المتبادل لضمان الاتصال الأمثل بين هذه الأشياء ، وبروتوكول التوجيه هو عنصر أساسي لهذا الغرض.

إن RPL هو أحد بروتوكولات التوجيه الرئيسية. طبيعة الأشياء والقيود الموضوعة مقنعة لاختيار نموذج الكشف عن الهجمات. و نظرًا لطبيعة شبكات RPL ، فمن الضروري تحديد وتحليل الهجمات التي يواجهها هذا البروتوكول.

في هذا العمل ، نقترح آلية لاكتشاف الهجمات (نظام كشف التسلل مينيرفا) التي تستهدف التوجيه في شبكة انترنت الاشياء والتي تعتمد على اكتشاف الاختراقات عبر السلوك بفضل التعلم الآلي.

لقد حددنا ، باستخدام محاكي Contiki Cooja ، عدة سيناريوهات للشبكة. بعد ذلك ، قمنا ببناء مجموعة البيانات الخاصة بنا باستخدام بيانات ومعايير مهمة لاكتشاف هجمات التوجيه في شبكة انترنت الأشياء ، وهو أمر ضروري لإنشاء نظام كشف التسلل (هجين ومتسلسل هرميا).

لقد اقترحنا نموذجًا من ثلاثة مستويات يجمع بين تقنيات التصنيف المختلفة (ثنائي ومتعدد التصنيف) مع التصفية لتقليص حجم مجموعة البيانات وتقليل الحد الأقصى من الإنذارات الكاذبة.

كانت النتائج التي تم الحصول عليها مرضية للغاية لأننا نجحنا في الحصول على معدل أدنى من الإنذارات الكاذبة (إيجابية كاذبة) ومعدل أقصى لاكتشاف الهجوم ، مع تقليص وقت التعلم والتنبؤ.

الكلمات المفتاحية : انترنت الأشياء، توجيه ، نظام كشف التسلل ، DODAG، RPL ، الأمن، الهجوم...

Table des matières

Liste des tableaux

Liste des figures

Liste des abréviations

6LoPAN: IPv6 over Low -Power Wireless Personal Area Networks

ACK: Acquittement

DAO: Dodag advertisement Object

DIO: Dodag information Object

DIS: Dodag information sollicitation

DODAG: Destination Oriented Directed Acyclic Graph

IDO: Internet Des Objects

IDS: Intrusion Detection Systems

IOT: Internet of Things

LTE-A : Long Term Evolution—Advanced

M2M: Machine to machine

P2P : Peer-to-Peer

RFID: Radio-frequency identification

RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks

WLAN: Wireless Local Area Network

WSN: Wireless sensor network

Introduction générale :

L’internet est devenu une nécessité dans la vie moderne, pour cette raison, les chercheurs ont essayé d'améliorer la vie quotidienne de l’être humain où il n’aura plus besoin d’intervenir pour fournir des services.

L'Internet des objets ou (IdO) ou en anglais Internet of Things (IoT) (Nous avons utilisé deux significations alternativement) n'est plus un rêve de science-fiction où les avancées technologiques en cours annoncent la connexion des objets quotidiens à l'internet , toutefois, des solutions ouvertes et interopérables doivent être utilisées pour assurer une communication optimale entre ces objets.

Aujourd’hui les progrès technologiques permettent maintenant d’envisager la connexion des objets du quotidien à l’Internet sans l’intervention humaine.

Le protocole de routage est un élément clé de cet objectif, car il permet à chaque objet de décider comment atteindre un autre objet, pour cela les contraintes qui s’appliquent aux objets (faible puissance, communications instables) doivent être prises en compte pour le développement de protocole de routages adaptés tel que : RPL, 6LowPAN, IEEE 802.15.4, et EPCglobal.

Néanmoins, l’IOT n’est qu’en ses début, plusieurs progrès restent à faire en matière de standardisations, de routage et d’identification, d’optimisation de la consommation d’énergie et surtout de sécurité, cette dernière reste un des problèmes majeurs dans le routage du réseau IdO et qui freine le déploiement rapide de cette technologie dû au changement fréquent de la nature des attaques ciblant les protocoles de routages.

La mise en œuvre de mesures de sécurité est essentielle pour garantir la confiance de ces réseaux auxquels sont connectés des dispositifs IdO, ce qui devenu ensuite un objectif pour les chercheurs pour trouver un mécanisme fiable de sécurité malgré les problèmes liés à la façon de distinguer les attaques qui peuvent passer inaperçues (faux négatifs) durant la phase d’apprentissage, ou aux fausses alertes lieu (faux positif).

C’est ainsi qu’apparaitre la nécessité des systèmes de détection d’intrusion (IDS) pour un environnement IdO dans l'analyse et l'interprétation des paquets reçus dans le réseau et détecter toute entrée non autorisée ou toute activité malveillante.

Dans ce contexte, notre travail consiste à élaborer un IDS et tester son efficacité en suivant un modèle (hybride et hiérarchique) pour la prise de décision de cet IDS, et enfin d'examiner son exactitude de la détection qui doit être proche de la certitude et minimiser le taux des fausses alertes.

Ce mémoire s'articule autour de cinq chapitres répartis comme suit :

- Le premier est consacré à la présentation des problèmes de sécurité informatique comme entrée dans le monde des menaces et de la vulnérabilité …,
- Le deuxième chapitre présente les IDS.
- Ensuite le troisième chapitre on expose une étude sur l'environnement IdO et les menaces liées aux messages de contrôles émis par les nœuds lors d'un routage RPL.
- Le quatrième chapitre présente notre contribution qui sert à réaliser un modèle IDS hybride pour la détection des véritables attaques de routage à partir de notre ensemble de données construit.
- Enfin Le dernier chapitre est consacré à la simulation et l'analyse des résultats obtenus de notre modèle proposé.

Chapitre1 : La Sécurité Informatique

Chapitre1 : La Sécurité Informatique

Introduction

En informatique, le terme sécurité recouvre tout ce qui concerne la protection des informations, De nos jours le monde est de plus en plus connecté et son système d'information est accessible de l'extérieur.

Ce chapitre à pour but d'éclaircir les notions fondamentales liées à la sécurité informatique tels que les définitions, les caractéristiques, les architectures et les domaines d'application.

I - Sécurité Informatique

La sécurité informatique est l'ensemble des moyens mis en œuvre pour minimiser la vulnérabilité d'un système contre des menaces accidentelles ou intentionnelles [1].

II - Services de la sécurité :

Toujours dans le domaine de l'informatique, le terme sécurité recouvre tout ce qui concerne la protection des informations.

Trois grands concepts ont été définis :

- Les fonctions de sécurité, qui sont déterminées par les actions pouvant compromettre la sécurité d'un établissement ;
- Les mécanismes de sécurité, qui définissent les algorithmes à mettre en œuvre ;
- Les services de sécurité, qui représentent les logiciels et les matériels mettant en œuvre des mécanismes dans le but de mettre à la disposition des utilisateurs les fonctions de sécurité dont ils ont besoin[2].

Assurer la sécurité revient alors à assurer les fonctions suivantes :

1 - Confidentialité

La *confidentialité* garantit aux utilisateurs qu'aucune donnée n'a pu être lue et exploitée par un tiers malveillant, les données (et l'objet et les acteurs) de la communication ne peuvent pas être connues d'un tiers non-autorisé.

2 - Authenticité

L'*authentification* consiste à demander à un utilisateur de prouver son identité, L'identité des acteurs de la communication est vérifiée.

3 - Intégrité

L'intégrité assure aux utilisateurs que leurs données n'ont pas été indûment modifiées et n'ont pas été altérées au cours de la transmission dans le réseau.

4 - Non-répudiation

Les acteurs impliqués dans la communication ne peuvent nier y avoir participé, elle empêche un utilisateur de contredire la réalité d'un échange de données.

5 - Disponibilité

Les acteurs de la communication accèdent aux données dans de bonnes conditions. (Cousin,)

III - Menace sur les réseaux :

1 - Vulnérabilité :

Internet est une mine d'informations pour les entreprises et pour les utilisateurs. En naviguant sur Internet, on peut accéder à des millions de pages Web. A l'aide de moteurs de recherche, on peut obtenir des informations qui sont nécessaires pour le travail, au moment où on a besoin(Cousin,).

Le Web est une ressource indispensable pour la productivité des entreprises, mais il y a aussi des dangers (les virus, les vers, les cookies…), alors le Web montre une faiblesse.

Pour résoudre ce problème, il faut une discipline où les listes de pages Web sont bloquées ou à conseillées, mais dans l'évolution très rapide du Web aujourd'hui, comment peut-on faire ? (Cousin,).

IV - Attaque :

De nos jours la sécurité du réseau informatique est devenue indispensable face aux attaques qui se multiplient rapidement. Pour contrarier ces attaques, les systèmes de sécurité visent à prévenir de ces dernières et à corriger les vulnérabilités exploitées. Il est alors nécessaire d'établir l'identification des menaces potentielles et de connaître les différents

procédés des attaquants afin de sécuriser le réseau. C'est pourquoi nous allons dans un premier temps analyser ce que nous appellerons « l'anatomie d'une attaque », puis dans un second temps, nous caractériserons ces attaques avec ses différents types.

V - Motivation des attaques :

Les motivations des attaques peuvent être de différentes sortes :

- Obtenir un accès au système .
- Voler des informations, tels que des secrets industriels ou des propriétés intellectuelles .
- Glaner des informations personnelles sur un utilisateur .
- S'informer sur l'organisation (entreprise de l'utilisateur, etc.) .
- Troubler le bon fonctionnement d'un service .
- Utiliser le système de l'utilisateur comme « rebond » pour une attaque .
- Utiliser les ressources du système de l'utilisateur.

1 - Anatomie d'une attaque :

Fréquemment appelés « les 5 P » dans la littérature, ces cinq verbes anglophones constituent le squelette de toute attaque informatique : Probe, Penetrate, Persist, Propagate, Paralyze.

Observons le détail de chacune de ces étapes[3]:

1-1 - Probe :

Consiste en la collecte d'informations par le biais d'outils comme whois, Arin, DNS lookup. La collecte d'informations sur le système cible peut s'effectuer de plusieurs manières, comme par exemple un scan de ports grâce au programme NAP pour déterminer la version des logiciels utilisés, ou encore un scan de vulnérabilités à l'aide du programme Nessus.

1-2 - Penetrate :

Utilisation des informations récoltées pour pénétrer un réseau. Des techniques comme le brute force ou les attaques par dictionnaires peuvent être utilisées pour

outrepasser les protections par mot de passe. Une autre alternative pour s'infiltrer dans un système est d'utiliser des failles applicatives que nous verrons ci-après.

1-3 - Persist :

Création d'un compte avec des droits de super utilisateur pour pouvoir se réinfiltrer ultérieurement. Une autre technique consiste à installer une application de contrôle à distance capable de résister à un reboot (ex : un cheval de Troie).

1-4 - Propagate :

Cette étape consiste à observer ce qui est accessible et disponible sur le réseau local.

1-5 - Paralyze :

Cette étape peut consister en plusieurs actions. Le pirate peut utiliser le serveur pour mener une attaque sur une autre machine, détruire des données ou encore endommager le système d'exploitation dans le but de planter le serveur.

Après ces cinq étapes, le pirate peut éventuellement tenter d'effacer ses traces, bien que cela ne soit rarement utile.

2 - Différents types d'attaques :

De nombreux types d'attaques du réseau ont été identifiés. Ces attaques sont généralement classées en trois principales catégories [4] :

- Les attaques dans le but de découvrir des informations.
- Les attaques par intrusions sont menées afin d'exploiter les faiblesses de certaines zones du réseau telles que les services d'authentification.
- Les attaques d'interruption de service (ou déni de service) saturent l'accès à une partie ou à l'intégralité d'un système. Les attaques d'interruption de service distribué (*DDOS : Distributed Denial of Service*) qui consistent à saturer ainsi plusieurs machines ou hôtes, sont encore plus nuisibles.
- Attaques passives : les attaques passives sont la *capture* du contenu d'un message et *l'analyse de trafic*. Elles sont très difficiles à détecter car elles ne causent aucune altération des données. Le but de l'adversaire est d'obtenir une information qui a été transmise figure 1 (Rhouma,) .

- Attaques actives : ces attaques impliquent certaines modifications du flot de données ou la création d'un flot frauduleux ; elles peuvent être subdivisées en 4 catégories : mascarade, rejeu, modification de messages et déni de service.
    - Une mascarade a lieu lorsqu'une entité prétend être une autre entité. Une attaque de ce type inclut habituellement une des autres formes d'attaque active.
    - Le rejeu implique la capture passive de données et leur retransmission ultérieure en vue de produire un effet non autorisé.
    - La modification des messages (man in the middle) signifie que certaines portions d'un message légitime sont altérées ou que les messages sont réorganisés.
    - Dénis de services [6]: d'une manière générale, l'attaque par déni de service (Denial of Service DoS) vise à rendre une application informatique incapable de répondre aux requêtes de ses utilisateurs par saturation de ses ressources. La Figure 1 montre un exemple de DoS qui s'appelle « ICMPFlood ».

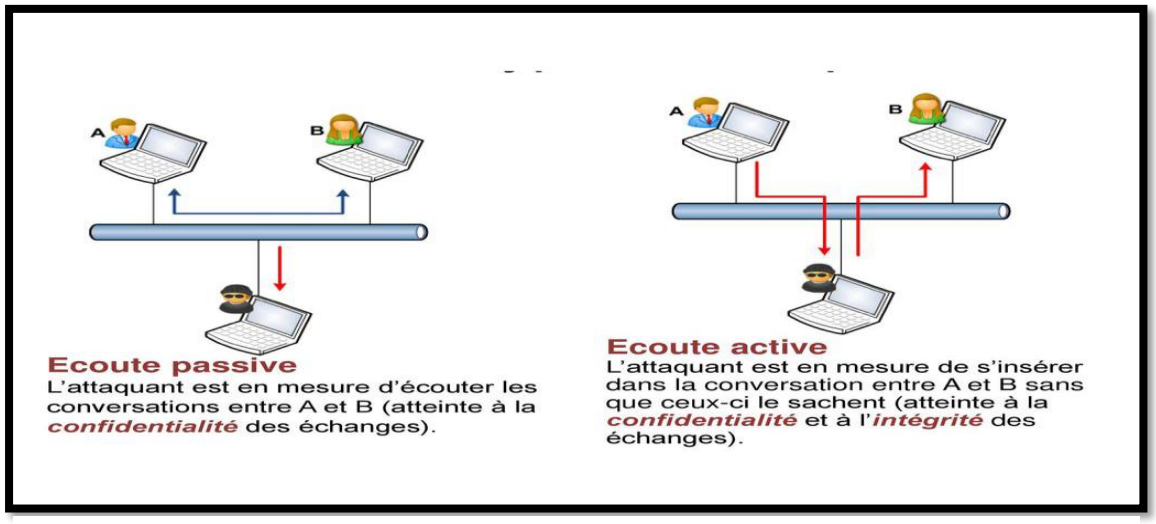


Figure 1: Ecoute passives et Ecoute actives.

3 - Buts des attaques

les attaques sur le système informatique ou réseau peut être largement classé selon leur but comme suit : interruption, interception, modification et fabrication.[7]

- **Interruption :** vise la **disponibilité** des informations
- **Interception :** vise la **confidentialité** des informations

- **Modification** : vise **l'intégrité** des informations
- **Fabrication :** vise **l'authenticité** des informations

4 - Différentes étapes d'une attaque :

Une attaque est l'exploitation d'une faille d'un système informatique connecté à un réseau. Pour réussir leur exploit, les attaquants tentent d'appliquer un plan d'attaque bien précis pour aboutir à des objectifs distincts.

La plupart des attaques, de la plus simple à la plus complexe fonctionnent suivant le même schéma [8] :

4-1 - Identification de la cible :

Cette étape est indispensable à toute attaque organisée, elle permet de récolter un maximum de renseignements sur la cible en utilisant des informations publiques et sans engager d'actions hostiles. On peut citer par exemple l'interrogation des serveurs DNS.

4-2 - Scanning :

L'objectif est de compléter les informations réunies sur une cible visée. Il est ainsi possible d'obtenir les adresses IP utilisés, les services accessibles de même qu'un grand nombre d'informations de topologie détaillée.

4-3 - Exploitation :

Cette étape permet à partir des informations recueillies d'exploiter les failles identifiées sur les éléments de la cible, que ce soit au niveau protocolaire, des services et applications ou des systèmes d'exploitation présents sur le réseau.

4-4 - Progression :

Il est temps pour l'attaquant de réaliser son objectif. Le but ultime étant d'obtenir les droits de l'utilisateur root sur un système afin de pouvoir y faire tout ce qu'il souhaite.

VI - Autres attaques réputées :

Attaque par identification des systèmes réseau :

1 - Le balayage de ports : [9]

1. Attaque par balayage ICMP
2. Attaque par balayage TCP

3. Attaque par balayage semi-ouvert TCP

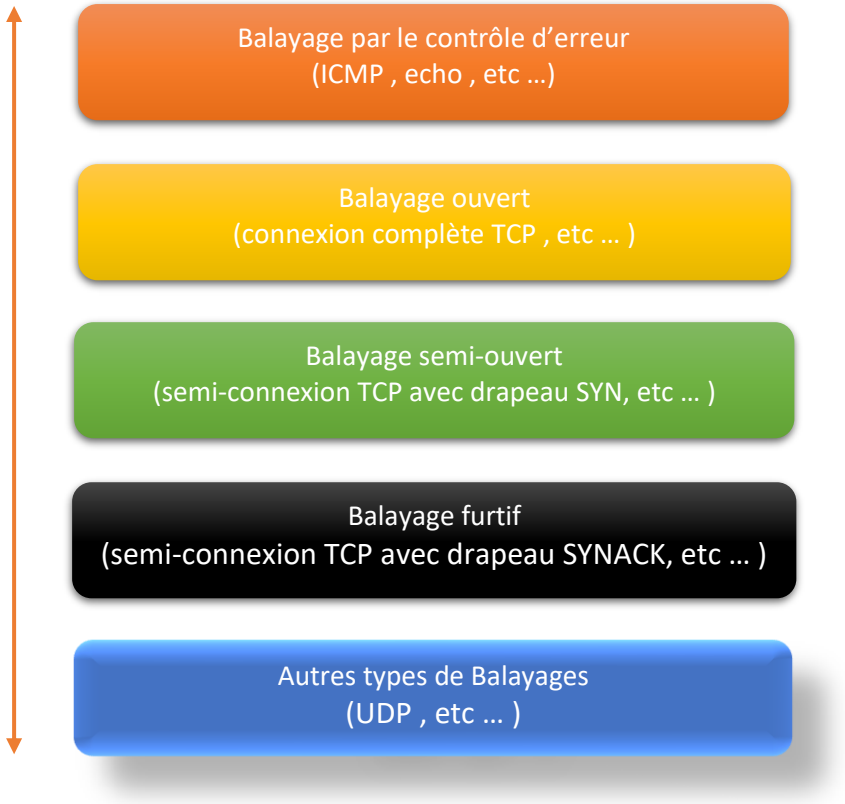


Figure 2 : Les différents types de balayages

VII - Types de logiciels malveillants :

1 - Virus :

Un virus est un logiciel capable de s'installer sur un ordinateur à l'insu de son utilisateur légitime. Le terme virus est réservé aux logiciels qui se comportent ainsi avec un but malveillant, parce qu'il existe des usages légitimes de cette technique dite de code mobile.

En général, pour infecter un système, un virus agit de la façon suivante : il se présente sous la forme de quelques lignes de code en langage machine binaire qui se greffent sur un programme utilisé sur le système cible, afin d'en modifier le comportement. Le virus peut être tout entier contenu dans ce greffon, ou il peut s'agir d'une simple amorce, dont le rôle va être de télécharger un programme plus important qui sera le vrai virus.

Une fois implanté sur son programme-hôte, le greffon possède aussi en général la capacité de se recopier sur d'autres programmes, ce qui accroît la virulence de l'infection et peut contaminer tout le système ; la désinfection n'en sera que plus laborieuse[10].

2 - Vers :

Un ver *(Worm)* est une variété de virus qui se propage par le réseau. Il se reproduit en s'envoyant à travers un réseau (e-mail, Bluetooth, chat.). Le ver contrairement aux virus, n'a pas besoin de l'interaction humaine pour pouvoir se proliférer.

3 - Cheval de Troie :

Un cheval de Troie *(Trojan horse)* est un logiciel qui se présente sous un jour honnête, utile ou agréable, et qui une fois installé sur un ordinateur y effectue des actions cachées et pernicieuses.

La différence essentielle entre un trojan et un ver réside dans le fait que le ver tente de se multiplier, Ce que ne fait pas le trojan.

4 - Porte dérobée :

Une porte dérobée *(backdoor)* est un logiciel de communication caché, installé par exemple par un virus ou par un cheval de Troie, qui donne à un agresseur extérieur accès à l'ordinateur victime, par le réseau [10].

5 - Bombe logique :

Une bombe logique est une fonction, cachée dans un programme en apparence honnête, utile ou agréable, qui se déclenchera à retardement, lorsque sera atteinte une certaine date, ou lorsque surviendra un certain événement. Cette fonction produira alors des actions indésirées, voir nuisibles [10].

Les bombes logiques présentent des caractéristiques similaires aux chevaux de Troie (incapacité de se reproduire et de se propager).

6 - Logiciel espion :

Un logiciel espion, comme son nom l'indique, collecte à l'insu de l'utilisateur légitime des informations au sein du système où il est installé, et les communique à un agent extérieur, par exemple au moyen d'une porte dérobée.

Une variété particulièrement toxique de logiciel espion est le *keylogger* (espion dactylographique), qui enregistre fidèlement tout ce que l'utilisateur tape sur son clavier et le transmet à son honorable correspondant ; il capte ainsi notamment identifiants, mots de passe et codes secrets [10].

7 - Spam :

Le spam est du courrier électronique non sollicité envoyé à un très grand nombre de personnes sans leur accord préalable.

Les messages électroniques non sollicités contiennent généralement de la publicité.

8 - Spyware :

Un spyware (mouchard) est un programme capable en plus de sa fonction propre de collecter des données sur ses utilisateurs et de les transmettre via Internet. Les spywares sont parfois confondus avec les adwares, ces logiciels dont l'auteur se rémunère par l'affichage de bannières publicitaires mais sans recueillir ni transmettre d'informations. [11]

Le but des mouchards est de recueillir le plus d'information possible de l'utilisateur. Ces mouchards sont présents dans de nombreux « freewares » ou « sharewares » et ils s'installent lors des téléchargements à l'insu des utilisateurs.

9 - Cookies :

Les cookies ne représentent pas de menace directe pour votre ordinateur ou les données qui y sont placées. Cependant, ils sont vraiment une menace pour la confidentialité : un cookie permet à un site Web de conserver vos références et de suivre à la trace vos visites du site. C'est pourquoi, si vous préférez garder l'anonymat, vous devriez désactiver les cookies en utilisant les paramètres de sécurité de votre navigateur.

Un cookie est un petit fichier au format texte d'un maximum de 4 Ko, envoyé ("offert", comme un biscuit ?) par le serveur d'un site Web et enregistré sur votre disque dur par votre navigateur.

Un cookie n'étant pas exécutable, il ne peut contenir de virus.

VIII - Mécanismes de sécurité :

La sécurité vise à garantir la confidentialité, l'intégrité et la disponibilité des services. Il faut mettre en place des mécanismes pour s'assurer que seules les personnes autorisées ont accès à l'information et que le service est rendu correctement. Parmi ces mécanismes, on peut citer :

1 - Cryptage :

Cryptographie est une science mathématique dans laquelle on fait les études des méthodes permettant de transmettre des données de manière confidentielle.

Afin de protéger un message, on lui applique une transformation qui le rend incompréhensible ; c'est ce qu'on appelle le chiffrement, qui, à partir d'un texte clair, donne un texte chiffré ou cryptogramme. Inversement, le déchiffrement est l'action qui permet de reconstruire le texte en clair à partir du texte chiffré. Dans la cryptographie moderne, les transformations en question sont des fonctions mathématiques, appelées algorithmes cryptographiques, qui dépendent d'un paramètre appelé clef.

Dans les réseaux, pour contrer les vols d'informations dans la voie de transmission, on utilise les techniques de cryptographie pour chiffrer et déchiffrer les messages transmis. Il existe à l'heure actuelle deux grands principes de cryptage : le cryptage symétrique basé sur l'utilisation d'une clef privée et le cryptage asymétrique qui repose sur un codage à deux clefs, une privée et l'autre publique.

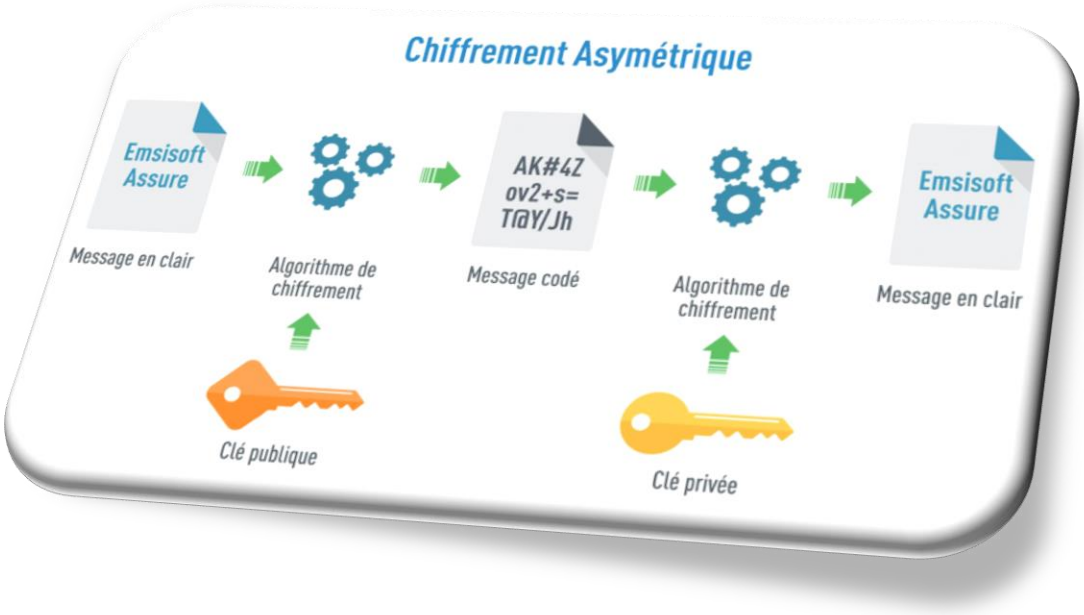


Figure 3:Chiffrement

2 - Pare-feu :

Un pare-feu (firewall) est une solution matérielle ou logicielle mise en place au sein de l'infrastructure du réseau afin de filtrer l'accès à des ressources réseau définies. Il ne laisse entrer que les utilisateurs autorisés, disposant d'une clef ou d'un badge, et crée une couche protectrice entre le réseau et le monde extérieur. Il est doté de filtres intégrés qui peuvent empêcher des documents non autorisés ou potentiellement dangereux d'accéder au système. Il enregistre également les tentatives d'intrusions dans un journal transmis aux administrateurs du réseau. Il permet également de contrôler l'accès aux applications et d'empêcher le détournement d'usage [5].

Le pare-feu permet à laisser passer tout ou partie des paquets qu'ils sont autorisés, et à bloquer et journaliser les échanges qui sont interdits.

La Figure 4 ci-dessous schématise le fonctionnement d'un pare-feu.

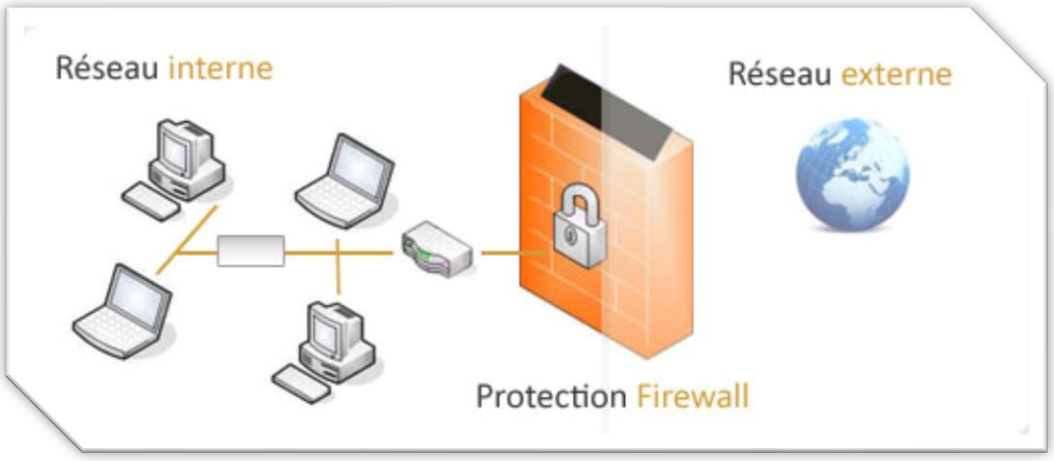


Figure 4 : Pare-feu

Le pare-feu est un IDS, mais il détecte seulement les attaques de l'extérieur. Pour Intranet, les pares-feux sont nécessaires, mais pas suffisants, pour commencer à implémenter une politique de sécurité.

Certains pares-feux laissent uniquement passer le courrier électronique. De cette manière, ils interdisent toute autre attaque qu'une attaque basée sur le service de courrier. D'autres pare-feu, moins strictes, bloquent uniquement les services reconnus comme étant des services

dangereux. Généralement, les pares-feux sont configurés pour protéger contre les accès non authentifiés du réseau externe.

3 - Antivirus :

Un antivirus est un logiciel qui protège une machine contre les virus. Les antivirus se fondent sur des fichiers de signatures et comparent alors les signatures génétiques du virus aux codes à vérifier. Certains programmes appliquent également la méthode heuristique tendant à découvrir un code malveillant par son comportement.

Les antivirus peuvent scanner le contenu d'un disque dur, mais également la mémoire de l'ordinateur. Pour les plus modernes, ils agissent en amont de la machine en scrutant les échanges de fichiers avec l'extérieur, aussi bien en flux montant que descendant. Ainsi, les courriers sont examinés, mais aussi les fichiers copiés sur ou à partir de supports amovibles tels que cédéroms, disquettes, connexions réseau, ...

Aujourd'hui, il y a beaucoup d'antivirus comme Norton Antivirus, McAfee Antivirus, Kaspersky Antivirus...

4 - VPN :

Les réseaux privés virtuels (VPN : Virtual Privat Network) permettent à l'utilisateur de créer un chemin virtuel sécurisé entre une source et une destination. Avec le développement d'Internet, il est intéressant de permettre ce processus de transfert de données sécurisé et fiable. Grâce à un principe de tunnel (tunneling) dont chaque extrémité est identifiée, les données transitent après avoir été chiffrées.

Un des grands intérêts des VPN est de réaliser des réseaux privés à moindre coût. En chiffrant les données, tout se passe exactement comme si la connexion se faisait en dehors d'I internet. Il faut par contre tenir compte de la toile, dans le sens où aucune qualité de service (QoS) n'est garantie.

Le principe du VPN est basé sur la technique du tunneling. Cela consiste à construire un chemin virtuel après avoir identifié l'émetteur et le destinataire. Ensuite la source chiffre les données et les achemine en empruntant ce chemin virtuel.

Les données à transmettre peuvent appartenir à un protocole différent d'IP. Dans ce cas le protocole de tunneling encapsule les données en rajoutant un entête. Permettant le routage

des trames dans le tunnel. Le tunneling est l'ensemble des processus d'encapsulation, de transmission et de dés encapsulation .[12]

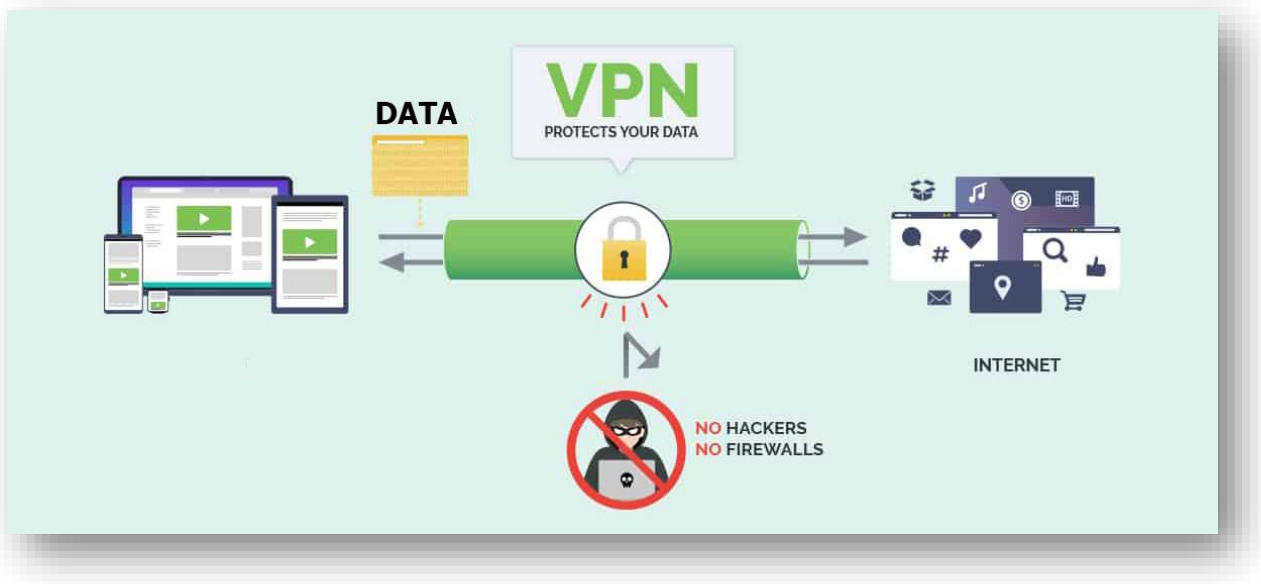


Figure 5 : VPN.

Conclusion

Dans ce chapitre nous avons vu comment un attaquant peut compromettre un système informatique en suivant une stratégie bien définie. Pour remédier à ces problèmes, des solutions de sécurité efficaces sont mises en œuvre par les administrateurs. Dans une optique d'optimisation de cette sécurisation, les systèmes de détection d'intrusions présentent un bon moyen de garantir cette sécurité des réseaux. C'est pourquoi nous entamerons dans le chapitre suivant l'étude des différents systèmes de détection d'intrusions et leur fonctionnement, ainsi que nous discutons aussi ses deux approches comportementales et par scénario.

Chapitre 2 : les systèmes de détection d'intrusion (IDS)

Chapitre 2 : les systèmes de détection d'intrusion (IDS)

Introduction :

L'informatique évolue, les systèmes et les réseaux informatiques deviennent de plus en plus complexes et les risques des failles augmentent donc la sécurité devienne de plus en plus difficile, car les délais laissés aux administrateurs sont souvent très courts.

Les attaques distribuées seront toujours redoutables si la plupart des machines ne sont pas protégées. Afin de détecter les attaques que peut subir un système, il est nécessaire d'avoir un logiciel spécialisé dont le rôle serait de surveiller les données qui transitent à travers ce système, et qui serait capable de réagir si des données semblant suspectes.

Plus communément appelé IDS, ces derniers se conviennent parfaitement pour réaliser cette tâche.

Ce chapitre a pour but d'éclaircir les notions fondamentales liées au IDS, par la suite, nous présentons une classification des IDS selon deux approches, comportementale et par scénario, de ce fait, nous parlerons sur la préparation des IDS, et comment faire cela ?

I - Système de détection d'intrusion :

Le premier modèle de détection d'intrusion est développé en 1984 par Dorothy Denning et Peter Neuman, qui s'appuie sur des règles de l'approche comportementale.

Ce système appelé (Intrusion Detection Expert System), en 1988 Il est développé à un **IDS** (système de détection d'intrusion).(*IDS*,)

Ce dernier est un ensemble de composants logiciels et/ou matériels destiné à repérer des activités anormales ou suspectes sur la cible analysée, un réseau ou un hôte, son rôle est de surveiller les données qui transitent sur ce système. Il permet ainsi d'avoir une action d'intervention sur les risques d'intrusion. Afin de détecter les attaques que peut subir un système ou réseau informatique. (DABOUR & HADJI,)

II - Architecture type d'un IDS :

1 - Architecture de base d'un système de détection d'intrusion :

Plusieurs architectures ont été proposées pour décrire les différents éléments intervenants dans un système de détection d'intrusion. L'architecture la plus simple est composée de trois modules : la source de données, l'analyseur des données et le module des réponses .[15]

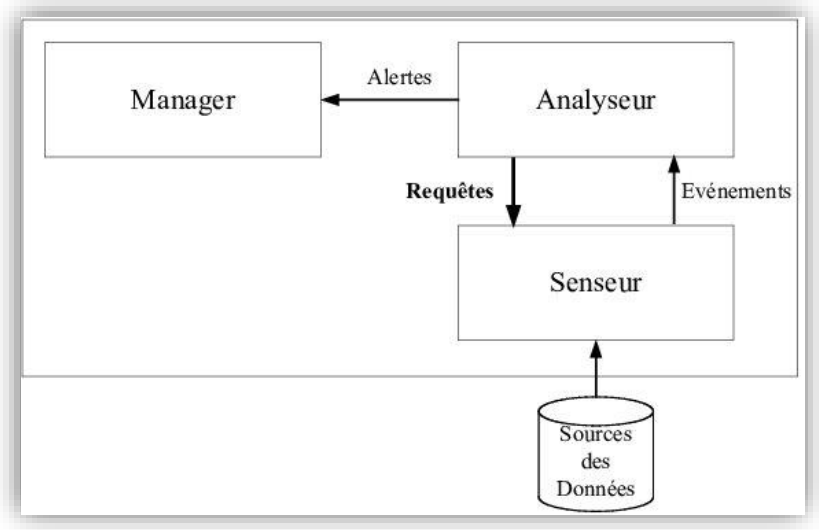


Figure 6 :L'architecture la plus simple d'un IDS [15]

1–1 – La source de données (Senseur):

appelée aussi sonde de capture ou senseur, elle s'occupe de la récupération des informations et des événements liées à la détection, pour les envoyer au module d'analyse. La position de la sonde de capture joue un rôle très important dans la qualité de la détection. Plusieurs sondes peuvent être utilisées dans le même IDS.

elles seront positionnées dans des points stratégiques du système.

1–2 – L'analyseur des données (Analyseur) :

C'est le cœur de l'IDS, ce module permet d'analyser les informations collectées par les sondes de capture. Il utilise une base de connaissances liée aux attaques, et pour la recherche des traces des activités malveillantes il applique des modèles d'analyse. Plusieurs modèles d'analyses existent et seront détaillés par la suite.

1–3 – Le module de réponses (Alertes) :

C'est le module qui assure les réponses de l'IDS aux activités malveillantes détectées. Les réponses peuvent être actives ou passives, c'est les contre-mesures nécessaires

pour empêcher les intrusions. Ça peut être un simple message d'alerte, une sauvegarde dans un fichier log ou bien interrompre une connexion.

Ces trois modules sont communs à la majorité des architectures proposées dans la littérature.

2 - L'architecture CIDF :

Le projet CIDF (Common Intrusion Detection Framework) , a visé le développement des protocoles et des interfaces de programmation d'application, pour permettre le partage de l'information et des ressources entre les projets de recherche. En effet, les modules dans le CIDF échangent des données dans un format standard, qui est basé sur un langage de communication spécifique : Common Intrusion Specification Language (CISL) . En plus, l'architecture CIDF permet la réutilisation des composants d'IDS déjà développés(*CIDF*,).

L'architecture CIDF, utilise quatre modules : Générateur d'événement, Analyseur d'événement, Unités de réponse et une Bases de données d'événements (figure 10). Les trois premiers modules jouent les mêmes rôles que ceux cités dans la section précédente. Tandis que la Base de données des événements est utilisée pour le stockage des évènements et des données analysées(*CIDF*,).

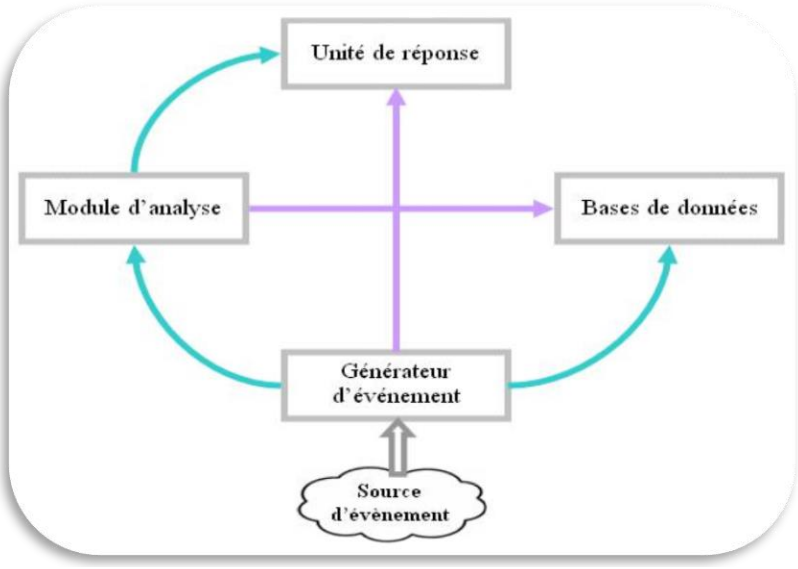


Figure 7 :L'architecture CIDF [79]

3 - L'architecture IDWG :

Dans l'architecture proposée par le groupe IDWG (Intrusion Detection exchange format Working Group) de l'IETF(Internet Engineering Task Force) , on trouve les trois modules cités précédemment couplés avec d'autres composants (figure 11). Dans cette architecture, l'objectif été la définition d'un standard de communication entre les composants du système de détection d'intrusion. Cette architecture définit un format d'échange de message pour les IDS: *Intrusion Detection Message Exchange Format (*IDMEF*)*, qui contient implicitement un modèle de données[17].

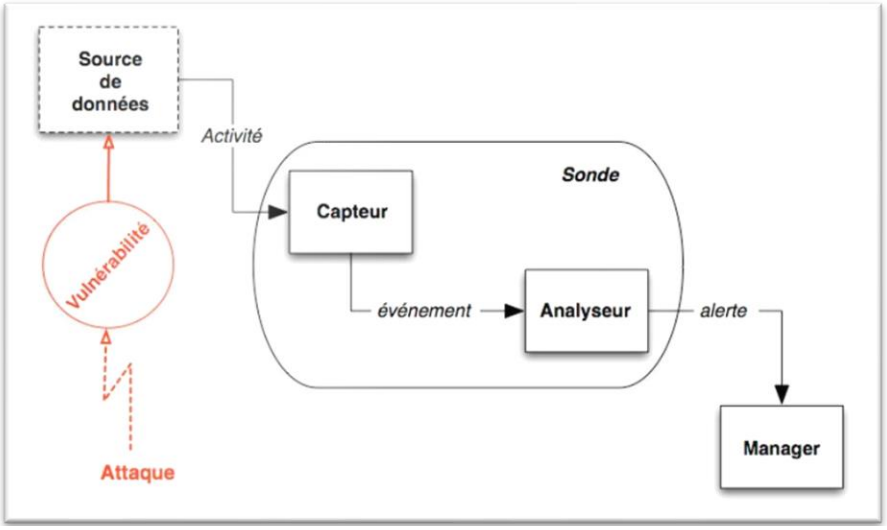


Figure 8:Détection d'intrusions: corrélation d'alertes [80]

Cette architecture est composée des modules suivants [17] :

3-1 - Source de données :

un dispositif qui génère de l'information sur les activités des entités du système d'information comme un analyseur réseau ou un système d'audit.

3-2 - Capteur :

un mécanisme de filtrage et de formatage de l'information brute provenant d'une source de données ; il génère des événements.

3-3 - Événement :

un message émis par un capteur. C'est l'unité élémentaire utilisée pour représenter une étape d'un scénario d'attaque connu. Ces événements sont parfois appelés événements d'audit, ou données d'audit.

3-4 - Analyseur :

un mécanisme d'analyse des événements à la recherche de traces d'intrusions.

3-5 - Alerte :

un message émis par un analyseur s'il trouve des traces d'intrusion.

3-6 - Sonde :

un ensemble constitué d'un capteur et d'un analyseur.

3-7 - Manager :

un composant permettant à l'administrateur du système de configurer les différents éléments (capteur, analyseur) et de gérer les alertes reçues.

III - Normalisation dans le domaine de la détection d'intrusion

De nos jours, il existe deux stratégies prioritaires en matière de normalisation dans le domaine de la détection d'intrusions.

La première ligne de recherche est la **création de protocoles et d'interfaces** qui permettent d'organiser la communication entre les systèmes de détection d'intrusion de différents fabricants.

Le second est le **développement des exigences** pour tester et certifier le système de détection d'intrusion.[18]

La première étape dans cette direction est le développement de la norme commune de cadre de détection d'intrusion (CIDF).

La création du groupe de travail sur la détection des intrusions (IDWG), sous l'égide de l'IETF, a servi comme prolongement de cette stratégie.

Actuellement, l'IDWG est sur le point d'accomplir ses travaux sur la détermination des exigences qui permettront la coordination et l'intégration du fonctionnement des systèmes de détection d'intrusion de différents fabricants.

IV - Types de système détection d'intrusion :

Les différents IDS se caractérisent par leur domaine de surveillance. Il existe trois grandes familles distinctes d'IDS :

1 - La détection d'intrusion basée sur l'hôte :

L'HIDS (Host Based IDS) surveille le trafic sur une seule machine. Il analyse les journaux systèmes, les appels, et enfin vérifie l'intégrité des fichiers. Un HIDS a besoin d'un système sain pour vérifier l'intégrité des données. Si le système a été compromis par un pirate, le HIDS ne sera plus efficace .(*Les sondes de sécurité IDS/IPS,*)

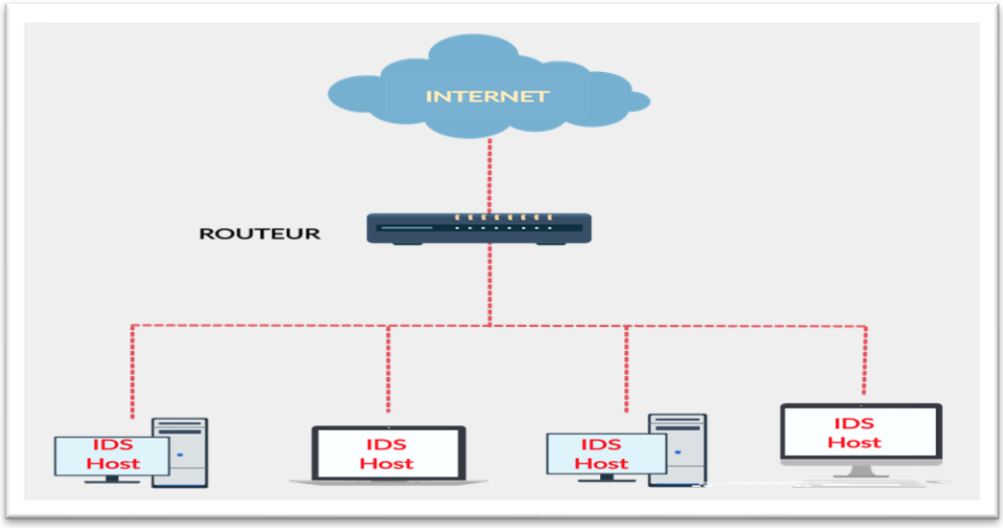


Figure 9 : Exemple d'une architecture d'un HIDS [14]

2 - La détection d'intrusion réseau NIDS :

Les NIDS sont des IDS utilisés pour protéger un réseau. Ils comportent généralement comme une sonde (machine par exemple) qui écoute et surveille en temps réel tout le trafic réseau, puis analyse et génère des alertes s'il détecte des intrusions ou des paquets semblent dangereux.(*Les sondes de sécurité IDS/IPS,*)

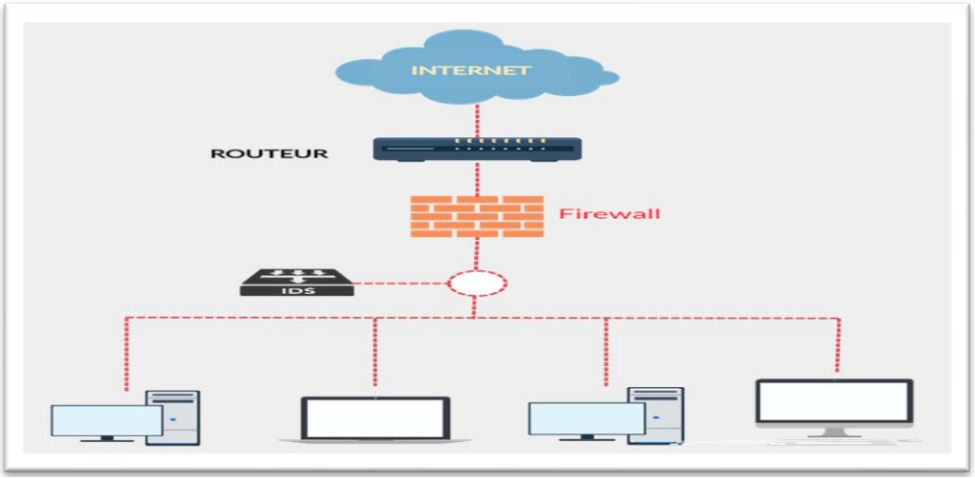


Figure 10 :Exemple d'une architecture d'un NIDS [14]

3 - Système de détection d'intrusion Hybride

Les IDS hybrides rassemblent les caractéristiques des NIDS et HIDS. Ils permettent, de surveiller le réseau et les terminaux. Les sondes sont placées en des points stratégiques, et agissent comme NIDS et/ou HIDS suivant leurs emplacements. Toutes ces sondes remontent alors les alertes à une machine qui va centraliser le tout, et lier les informations d'origines multiples. Ainsi, on comprend que les IDS hybrides sont basés sur une architecture distribuée, ou chaque composant unifie son format d'envoi. Cela permet de communiquer et d'extraire des alertes plus exactes.(*Les sondes de sécurité IDS/IPS*,)

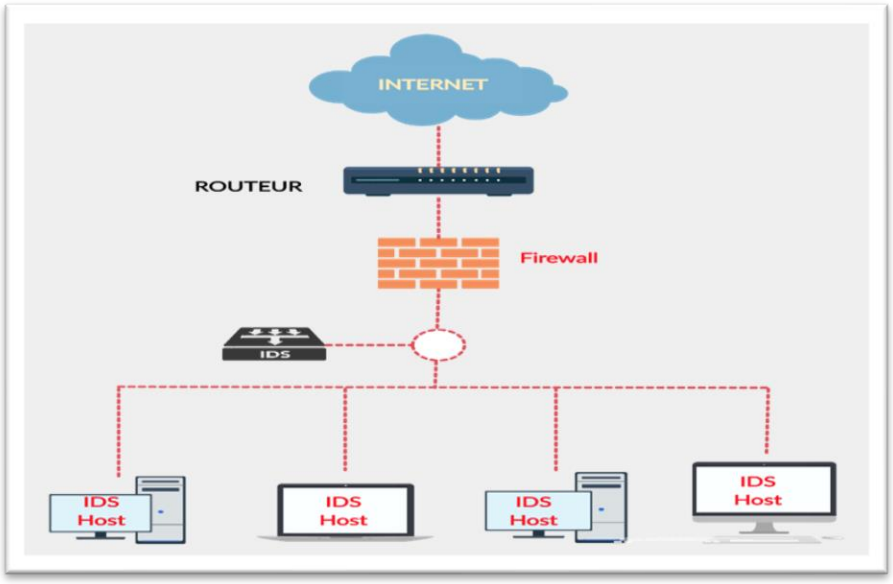


Figure 11 :Exemple d'une architecture d' Hybride [14]

V - Comparaison entre les types d'IDS :

type	Avantages	Inconvénients
NIDS	-Les capteurs peuvent être bien sécurisés puisqu'ils se contentent d'observer le trafic. -Détecter plus facilement les scans grâce aux signatures. -Filtrage de trafic. -assurer la sécurité contre les attaques puisqu'il est invisible.	-La probabilité de faux négatifs (attaques non détectées) est élevée et il est difficile de contrôler le réseau entier. -Ils doivent principalement fonctionner de manière cryptée d'où une complication de l'analyse des paquets. -A l'opposé des IDS basés sur l'hôte, ils ne voient pas les impacts d'une attaque.

HIDS	-Découvrir plus facilement un Cheval de Troie puisque les informations et les possibilités sont très étendues. -Détecter des attaques impossibles à détecter avec des IDS réseau puisque le trafic est souvent crypté. -Observer les activités sur l'hôte avec précision.	-Ils ont moins de facilité à détecter les scans. -Ils sont plus vulnérables aux attaques de type DoS. -Ils consomment beaucoup de ressources CPU.
Hybrides	-moins de faux positifs. -meilleure corrélation (la corrélation permet de générer de nouvelles alertes à partir de celles existantes). -possibilité de réaction sur les analyseurs.	-taux élevé de faux positifs.

Tableau 1 :la comparaison entre types d'IDS [19]

VI - Classification des systèmes de détection d'intrusion :

Les différents systèmes de détection d'intrusion disponibles peuvent être classés selon plusieurs critères qui sont : [20]

- La méthode de détection.
- Le comportement du système après la détection.
- La source des données.
- La fréquence d'utilisation.
- La figure 14 ci-dessous illustre les détaille de chaque critère.

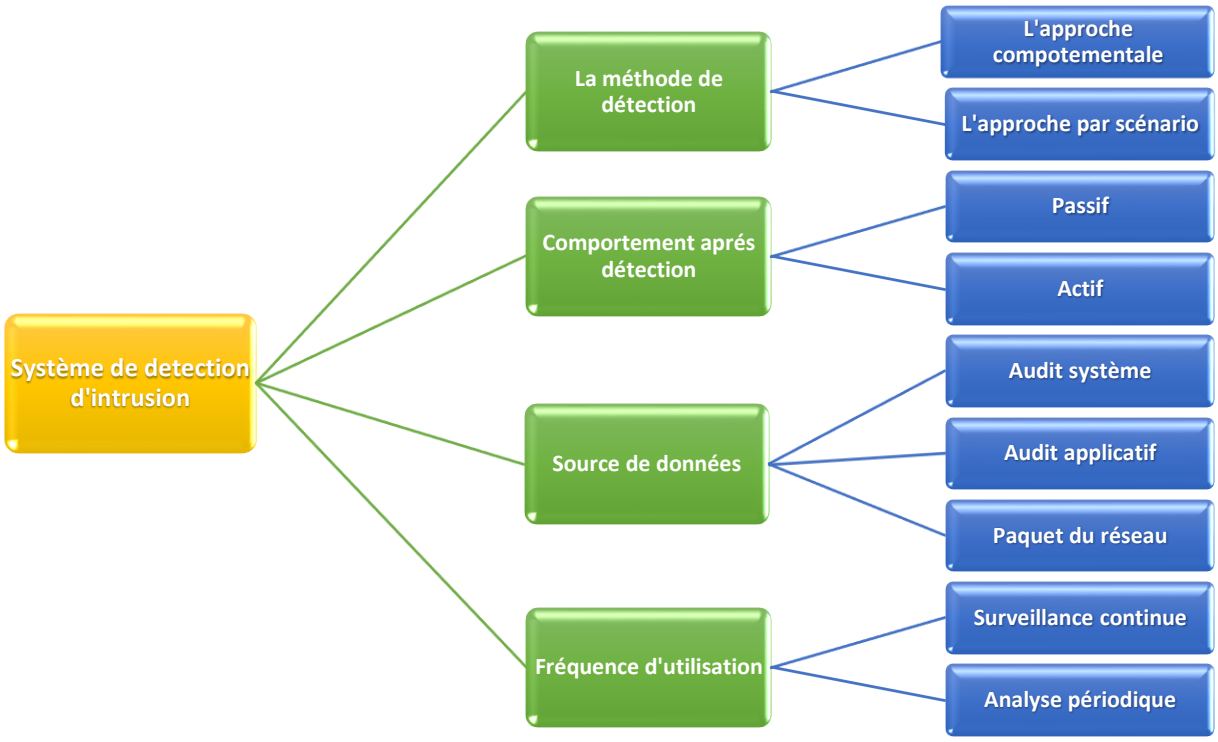


Figure 12:classification d'un système de détection d'intrusion [20].

1 - Méthodes de détection des IDS :

les techniques de détection d'intrusions se répartissent en deux classes :

détection **d'anomalies**, aussi appelée **approche comportementale**, et détection par **signature**, dite **détection de mauvais usage**, **détection par l'apparence** ou encore **approche par scénario** .[21]

1-1 - Approche par scénario ou par signature :

Cette technique s'appuie sur les connaissances des techniques utilisées par les attaquants contenues dans la base de donnée, elle compare l'activité de l'utilisateur à partir de la base de donnée, ensuite elle déclenche une alerte lorsque des événements hors profil se produisent.[22]

1-2 - L'approche comportementale :

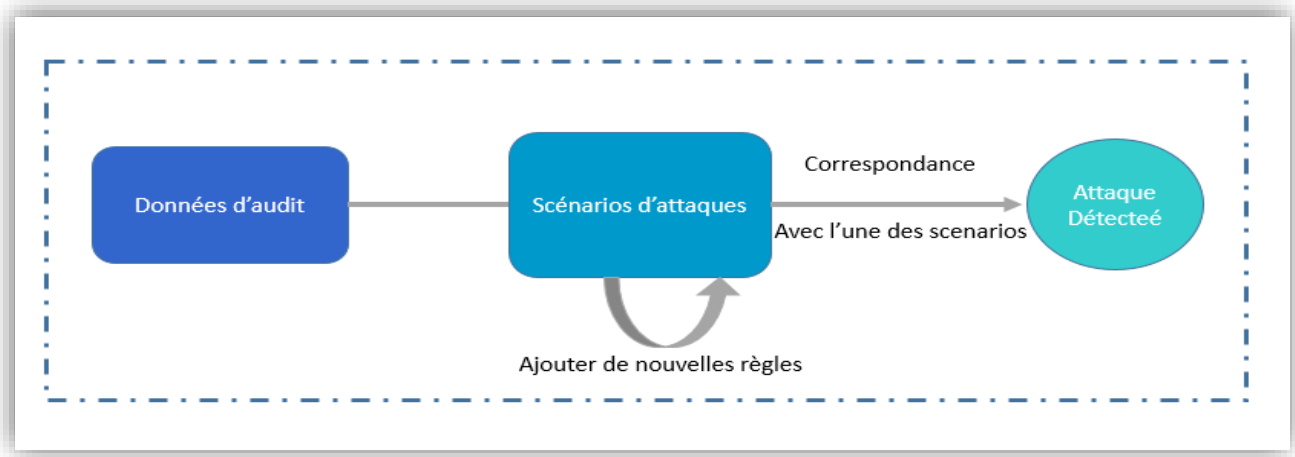


Figure 13 :Fonctionnement d'un IDS par l'approche basée connaissance [22]

Cette technique consiste à détecter une intrusion en fonction du comportement de l'utilisateur ou d'une application, autrement dit c'est créer un modèle basé sur le comportement habituel du système et surveiller toute déviation de ce comportement.

Plusieurs paramètres sont possibles : la charge CPU, le volume de données échangées, la durée et l'heure de connexion sur des ressources, la répartition statistique des protocoles et applications utilisés...etc . [23]

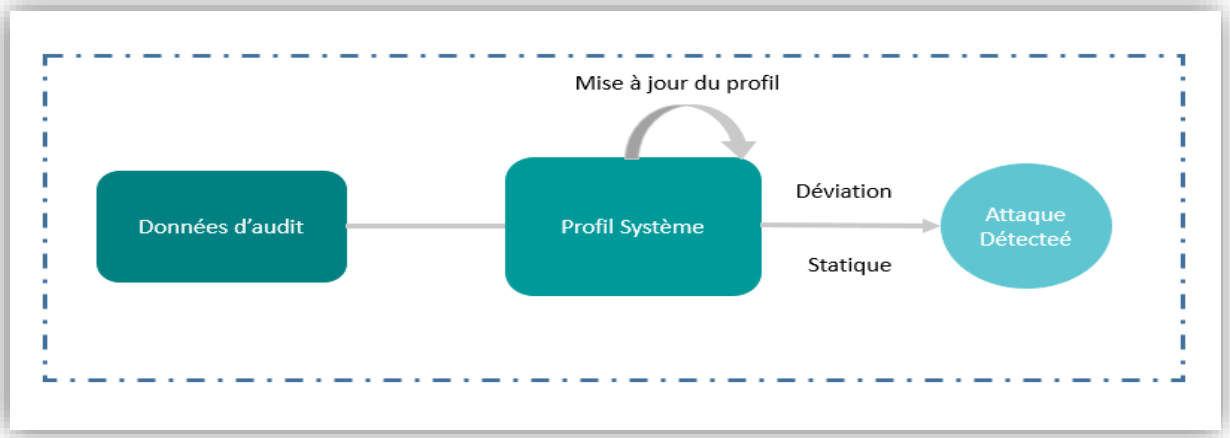


Figure 14:Fonctionnement d'un IDS par l'approche comportementale [23]

L'idée principale de cette approche est de considérer toute déviation, toute anomalie dans le comportement comme une intrusion. Cette hypothèse est certainement fausse : des événements ou des comportements rares peuvent tout a fait être légitimes du point de vue de la politique de sécurité du système.

Le système est susceptible d'émettre des faux positifs. Tant que le nombre de faux positifs reste suffisamment faible, la méthode peut être valide.

Cela conduit à poser deux questions essentielles, dans le domaine de la détection d'intrusions comportementale, sur le caractère correct et complet du modèle de comportement normal (voir la figure 17).[23]

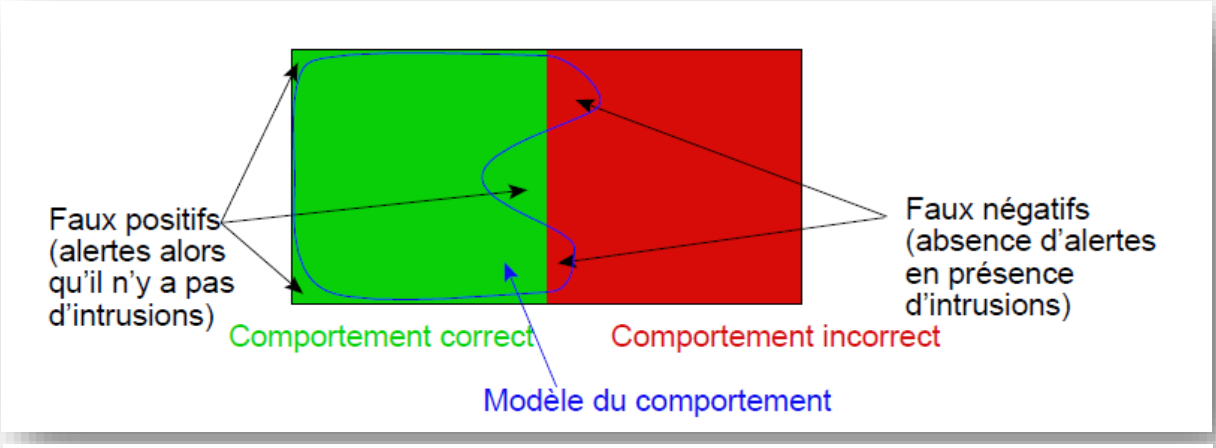


Figure 15 :Caractères complet et correct du modèle de comportement normal [23]

1-3 - Comparaison entre les deux approches :

Le tableau 2 suivant établi une comparaison entre les caractéristiques des deux précédentes approches [23]:

Scénario	Comportementale
Pas de faux positifs	Faux positifs nombreux
Pas de détection d'attaques non connues	Prise en compte de nouvelles attaques
Mise à jour rapide	Mise à jour délicate (phase d'entrainement)

Tableau 2 :Comparaison entre l'approche par scénario et l'approche

1-4 - Approche comportementale ou approche par scénarios ?:

Chacune de ces deux approches présente des avantages et des inconvénients (voir tableau 3). C'est pourquoi une approche hybride semble indispensable.(*Les systèmes de détections d'intrusions*,)

	Avantages	Inconvénients
Comportementale	Détection d'intrusion inconnue possible.	Choix délicat des mesures à retenir pour un système cible donné. Pour un utilisateur au comportement erratique, toute activité est "normale". En cas de profonde modification de l'environnement du système cible, déclenchement d'un flot ininterrompu d'alarmes (faux positifs) Utilisateur pouvant changer lentement de comportement dans le but d'habituer le système à un comportement intrusif (faux négatifs).
Par scénarios	Prise en compte des comportements exacts des attaquants potentiels.	Base de règles délicate à construire. Seules les attaques contenues dans la base sont détectées.

Tableau 3:Approche comportementale ou approche par scénarios ? [24]

2 - Comportement après la détection d'intrusion :

Il existe deux types de réponses, suivant les IDS utilisés. La réponse passive est disponible pour tous les IDS, la réponse active est plus ou moins implémentée.

2-1 - Réponse passive :

Lorsqu'une attaque est détectée, le système d'intrusion ne prend aucune action, il génère seulement une alarme en direction de l'administrateur système sous forme d'une alerte lisible qui contient les informations à propos de chaque attaque. Les réponses passives se traduisent la plupart du temps par des opérations de reconfiguration automatique d'un firewall afin de bloquer les adresses IP source impliquées dans les intrusions. Mais si le pirate prend une adresse

IP sensible telle qu'un routeur d'accès ou un serveur DNS, l'entreprise qui implémente une reconfiguration systématique d'un firewall risque tout simplement de se couper du monde extérieur.[22]

2-2 - Réponse active :

La réponse active consiste à répondre directement à une attaque, elle implique des actions automatisées prises par un IDS qui permet de couper rapidement une connexion suspecte quand le système détecte une intrusion. Par exemple interrompre le progrès d'une attaque pour bloquer ensuite l'accès suivant de l'attaquant.

Mais cela risque de se voir exposer à une contre-attaque par le pirate.

3 - La nature des données analysées :

La nature des données analysées sont composées de : [22]

3-1 - Les audits systèmes :

Les audits systèmes sont produits par le système d'exploitation d'un hôte. Ces données permettent à un IDS de contrôler les activités d'un utilisateur sur un hôte.

3-2 - Les audits applicatifs :

Les données à analyser sont produites directement par une application, par exemple des fichiers logs générés par les serveurs FTP et les serveurs Web. L'avantage de cette catégorie est que les données produites sont très synthétiques, elles sont riches et leur volume est modéré. Ces types d'informations sont généralement intégrés dans les IDS basés sur l'hôte.[22]

3-3 - Les sources d'informations réseau :

Ce sont des données du trafic réseau. Cette source d'informations est prometteuse car elle permet de rassembler et analyser les paquets de données circulant sur le réseau. Les IDS qui exploitent ces sources de données sont appelés : Les IDS basés réseau NIDS.[22]

4 - La fréquence d'utilisation :

La fréquence d'utilisation d'un système de détection d'intrusion peut exister selon deux formes :[22]

4-1 - Surveillance périodique :

Ce type de système de détection d'intrusion analyse périodiquement les différentes sources de données à la recherche d'une éventuelle intrusion ou une anomalie passée.

4-2 - Surveillance en temps réel :

Les systèmes de détection d'intrusions en temps réel fonctionnent sur le traitement et l'analyse continue des informations produites par les différentes sources de données. Elle limite les dégâts produits par une attaque car elle permet de prendre des mesures qui réduisent le progrès de l'attaque détectée.[22]

VII - Exemples d'application des systèmes de détection d'intrusion :

1 - Systèmes de détection d'intrusion réseaux :

- Cisco Secure IDS
- ISS RealSecure
- Symantec IDS
- Checkpoint SmartDefense
- Computer Associates eTrust
- Snort
- Bro
- Suricata

2 - Systèmes de détection d'intrusion hôtes :

- AIDE
- Chkrootkit
- DarkSpy
- IceSword .
- OSSEC
- Rootkit Unhooker
- Tripwire .

3 - Hybrides :

- Prelude et OSSIM

VIII - Domaines d'application :

1 - Systèmes distribués :

Les systèmes de détection et de prévention d'intrusions dans les ***systèmes distribués*** permettent de repérer et d'empêcher l'intrusion d'un utilisateur malveillant dans un système distribué comme une ***grille informatique*** ou un ***réseau en nuage (cloud)***.[25]

2 - Internet des objets :

Avec la constante augmentation des réseaux de capteurs, leur nombre devrait approcher les 26 milliards en 2020[26], l'***internet des objets*** représente de nombreux enjeux de sécurité, notamment dus à leur faible puissance de calcul, leur hétérogénéité, le nombre de capteurs dans le réseau ainsi que la ***topologie du réseau***[27]. De ce fait, les systèmes de détection d'intrusion traditionnels ne peuvent pas directement être appliqués aux réseaux de capteurs. Néanmoins, de nombreuses avancées ont été présentées au cours des années 2000-2010 pour pallier cette problématique.

IX - Critères de Choix D'un IDS :

Les systèmes de détection d'intrusion sont devenus indispensables lors de la mise en place d'une infrastructure de sécurité opérationnelle. Ils s'intègrent donc toujours dans un contexte et dans une architecture imposants des contraintes très diverses[22].

Certains critères imposant le choix d'un IDS peuvent être dégagés:

1 - Fiabilité :

Les alertes générées doivent être justifiées et aucune intrusion ne doit pouvoir lui échapper. Une intrusion non signalée constitue une défaillance de l'IDS, appelée **faux négatif.**

2 - Pertinence des alertes :

toute alerte doit correspondre à une intrusion effective, toute « fausse alerte » (appelée également **faux positif**) diminue la pertinence de l'IDS. (voir Figure 12), un IDS est parfaitement fiable en absence de faux négatif ; il est parfaitement pertinent en l'absence de faux positif.

3 - Réactivité :

Un IDS doit être capable de détecter les nouveaux types d'attaques le plus rapidement possible , pour cela il doit rester constamment à jour. Des capacités de mise à jour automatique sont indispensables.

4 - Facilité de mise en œuvre et adaptabilité :

Un IDS doit être facile à mettre en œuvre , surtout s'adapter au contexte dans lequel il doit opérer .

Il est inutile d'avoir un IDS émettant des alertes en moins de 10 secondes si les ressources nécessaires à une réaction ne sont pas disponibles pour agir dans les mêmes contraintes de temps.

5 - Performance :

la mise en place d'un IDS ne doit en aucun cas affecter les performances des systèmes surveillés[22].

De plus, il faut toujours avoir la certitude que l'IDS a la capacité de traiter toute l'information à sa disposition (par exemple un IDS réseau doit être capable de traiter l'ensemble du flux pouvant se présenter à un instant donné sans jamais supprimer de paquets) car dans le cas contraire il devient trivial de masquer les attaques en augmentant la quantité d'information.

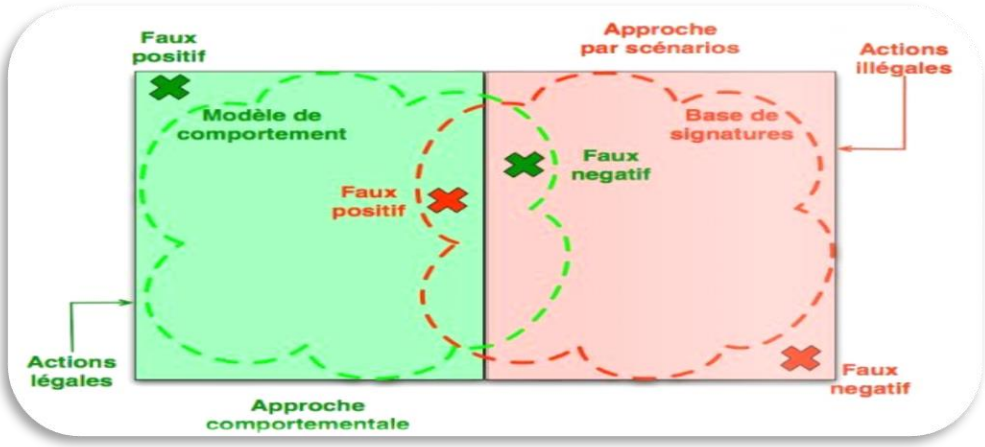


Figure 16 :Problèmes des IDS [22]

Les IDS proposent les fonctions suivantes[22]:

- Détection d'attaques (actives ou passives)
- Génération des rapports
- Outils de corrélation avec d'autres éléments de l'architecture de sécurité
- Réaction aux attaques par le blocage de route ou la fermeture de connexion

X - Choix du placement d'un IDS :

Le placement des IDS va dépendre de la politique de sécurité définie dans le réseau. Mais il existe des positions qu'on peut qualifier de standards, par exemple il serait intéressant de placer des IDS[28] :

- Dans la zone démilitarisée (attaques contre les systèmes publics) **.**
- Dans le (ou les) réseau(x) privé(s) (intrusions vers ou depuis le réseau interne) .
- Sur la patte extérieure du firewall (détection de signes d'attaques parmi tout le trafic entrant et sortant, avant que n'importe quelle protection intervienne).

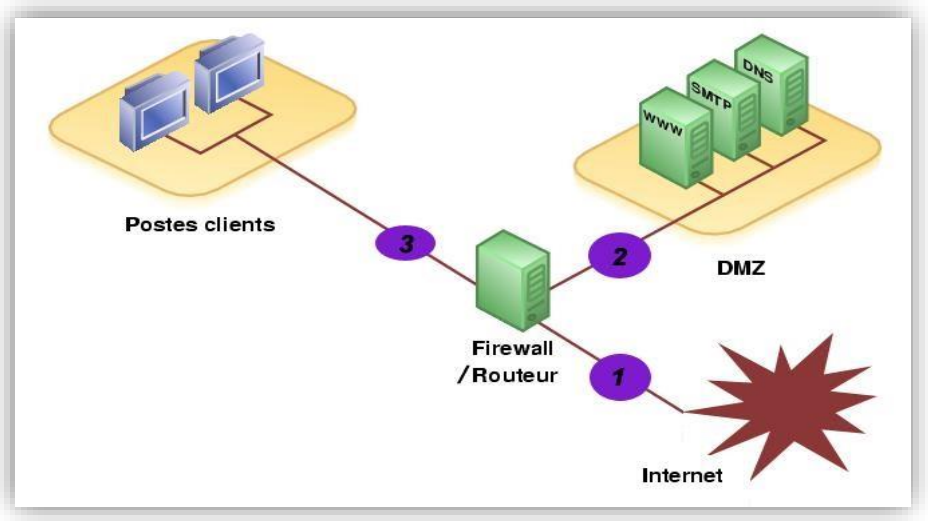


Figure 17:Choix du Placement d'un IDS [28]

Il est important de bien définir les zones sensibles du système (réseau), ainsi que les zones les plus attractives pour un pirate. Il faut aussi voir qu'au-delà de l'architecture du réseau, il faut prendre en compte l'organisation de la sécurité existante[28]:

- Recherche-t-on une administration centralisée ?

- Quel est l'existant organisationnel de la surveillance du réseau ?
- Quels sont les compétences et les moyens en internes pour gérer les IDS ?

XI - Les fonctions principales d'un IDS :

Donc un IDS a quatre fonctions principales(*Les systèmes de détection d'intrusions*,): l'analyse, la journalisation, la gestion et l'action.

1 - Analyse :

analyse des journaux du système pour identifier des intrusions dans la masse de données recueillie par l'IDS. Il y a deux méthodes d'analyse : une basée sur les signatures d'attaques et l'autre sur la d´détection d'anomalies (comportementale).

2 - Journalisation :

enregistrement des évènements dans un fichier de log. Exemple : d'évènement : arrivée d'un paquet, tentative de connexion.

3 - Gestion :

les IDS doivent être administres de manière permanente. On peut assimiler un IDS `a une camera de sécurité.

4 - Action :

alerter l'administrateur quand une attaque dangereuse est détectée.

XII - Limite des IDS :

Parmi les faiblesses des IDS on trouve :[30]

- Nombreux faux positifs.
- Configuration complexe et longue.
  - Nombreux faux positifs après configuration.
- Pas de connaissance de la plate-forme.
  - De ses vulnérabilités.
  - Du contexte métier.
- Les attaques applicatives sont difficilement détectables.
  - Injection SQL.
  - Exploitation de CGI mal conçus.
- Des évènements difficilement détectables.

- Scans lents / distribués
- Canaux cachés / tunnels.

- Pollution des IDS.
  - Consommation des ressources d'IDS.
  - Perte de paquets.
  - Déni de service contre IDS / opérateur.
  - Une attaque réelle peut passer inaperçue.
- Attaque contre IDS lui-même.
- Ils ne peuvent pas compenser les trous de sécurité dans les protocoles réseaux.
- Ils ne peuvent pas compenser des manques significatifs dans votre stratégie de sécurité, votre politique de sécurité ou votre architecture de sécurité.

XIII - Efficacité des systèmes de détection d'intrusions :

L'efficacité d'un système de détection d'intrusions est déterminée par les mesures suivantes : [20]

1 - Exactitude :

Le système de détection d'intrusions n'est pas exact s'il considère les actions légitimes des utilisateurs comme atypiques ou intrusives (faux positif).

2 - Performance :

Effectuer une détection en temps réel.

3 - Tolérance aux pannes :

Un système de détection d'intrusions doit être résistant aux attaques.

4 - Rapidité :

Un système de détection d'intrusions doit exécuter et propager son analyse d'une manière prompte pour permettre une réaction rapide dans le cas d'existence d'une attaque pour permettre à l'agent de sécurité de réagir.

5 - La complétude :

La complétude est la capacité d'un système de détection d'intrusion de détecter toutes les attaques .[31]

Conclusion :

Ce chapitre nous a permis de constater que les IDS sont de plus en plus fiables, d'où le fait qu'ils soient souvent intégrés dans les solutions modernes de sécurité. Les avantages qu'ils présentent par rapport aux autres outils de sécurité les favorisent. Il nous a également permis de comprendre que ces derniers sont indispensables aux entreprises afin d'assurer leur sécurité informatique.

Chapitre 3 : L'Internet des objets (IOT)

Chapitre 3 : L'Internet des objets (IOT)

Introduction :

L'internet des objets ou IdO (en anglais (the) Internet of Things ou IoT) est l'interconnexion entre l'Internet et des objets, des lieux et des environnements physiques. L'appellation désigne un nombre croissant d'objets connectés à l'Internet permettant ainsi une communication entre nos biens dits physiques et leurs existences numériques. Ces formes de connexions permettent de rassembler de nouvelles masses de données sur le réseau et donc, de nouvelles connaissances et formes de savoirs.

Dans ce chapitre, nous donnons une définition de l'Internet des objets et de ses notion bases, nous terminons par introduire des généralités sur la sécurité dans l'IOT avant d'entamer

Figure 18 : la distribution des appareils intelligents et une étude sur les personnes connectées [81]

particulièrement le sujet du routage IoT.

I - Définition :

L'Internet des objets (IoT) désigne l'interconnexion de millions d'appareils et de capteurs intelligents connectés à Internet. Ces capteurs et ces appareils connectés collectent et partagent des données qui seront utilisées et analysées par plusieurs organismes, dont des entreprises, des villes, des gouvernements, des hôpitaux et des particuliers.

Il faut savoir que **:** '' L'Internet des Objets est un réseau des réseaux qui permet, via des systèmes d'identification électronique normalisés et unifiés, et des dispositifs mobiles sans fil, d'identifier directement et sans ambiguïté des entités numériques et des objets physiques et ainsi de pouvoir récupérer, stocker, transférer et traiter, sans discontinuité entre les mondes physiques et virtuels, les données s'y rattachant [32]'.

II - Technologies de l'IoT :

L'IoT permet l'interconnexion des différents Objets intelligents via l'Internet. Ainsi, pour son fonctionnement, plusieurs systèmes technologiques sont nécessaires. "L'IoT désigne diverses solutions techniques (RFID, TCP/IP, technologies mobiles, etc.) qui permettent d'identifier des Objets, capter, stocker, traiter, et transférer des données dans les environnements physiques " [33]. En effet, bien qu'il existe plusieurs technologies utilisées dans le fonctionnement de l'IoT, nous mettons l'accent seulement sur quelques-unes qui sont, selon Han et Zhanghang, les technologies clés de l'IoT. Ces technologies sont les suivantes :

RFID, WSN et M2M, et elles sont définies comme suit :

- RFID : est une technologie sans fil qui est utilisée pour l'identification des Objets[34], elle englobe toutes les technologies qui utilisent des ondes radio pour identifier automatiquement des Objets ou des personnes. C'est une technologie qui permet de mémoriser et de récupérer des informations à distance grâce à une étiquette qui émet des ondes radio. Il s'agit d'une méthode utilisée pour transférer les données des étiquettes à des Objets, ou pour identifier ces Objets à distance. L'étiquette contient des informations stockées électroniquement pouvant être lues à distance [33].

- WSN : est un ensemble des nœuds qui communiquent sans fil et qui sont organisées en un réseau coopératif. Chaque nœud possède une capacité de traitement et peut contenir de différents types de mémoires, un émetteur-récepteur RF et une source

d'alimentation. Il peut aussi tenir compte des divers capteurs et actionneurs [35] Comme son nom l'indique, le WSN constitue un réseau de capteurs sans fil qui peut être une technologie nécessaire au fonctionnement de l'IoT.

- M2M : est l'association des technologies de l'information et de la communication avec des Objets intelligents dans le but de donner à ces derniers les moyens d'interagir sans intervention humaine avec le système d'information d'une organisation ou d'une entreprise [33]

III - Domaines D'applications :

Dans nos jours l'importance de l'internet des objets augmente jour par jour, les chercheurs estiment : ''que 3 millions de nouveaux terminaux se connecter à l'Internet chaque mois, dans les 4 prochaine année ce chiffre devrait atteindre les 30 milliards appareils connectées dans le monde entier ''.

L'utilisation de l'IOT permettra le développement de plusieurs applications intelligentes qui affecteront principalement les domaines abordés dans ce qui suit, avec un bref d'exemples de ses applications (Al-Fuqaha et al.) :

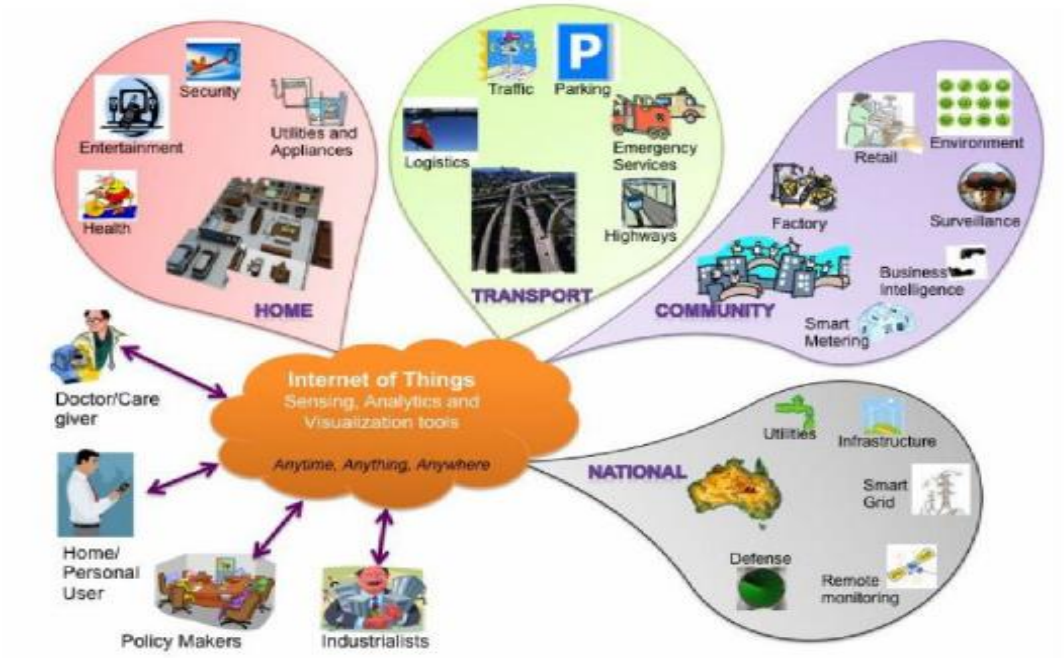


Figure 19 : Les domaines de l'internet des Objets

1 - L’internet des objets dans le domaine des sportifs :

De nombreux objets connectés comme des montres ou des bracelets connectés vous permettrez pendant la journée de calculer le nombre de pas effectuée, la distance parcourue, votre temps d’activités, les calories brulées, ainsi pendant la nuit en calculant vos heures de sommeil. Pour les passionnés du High-tech, c’est un grand marché qui s’ouvre pour eux ! De la montre connectée au téléviseur connecté en passant par les appareils photos, les drones,et les lunettes (Google glass)[33]..

2 - L’agriculture :

L'objectif principal de l'agriculture intelligente rendre fort la capacité des systèmes agricoles à prévoyaient les risques météorologiques, la maîtrise de l'irrigation, et la contribution de la sécurité alimentaire grâce à la collecte et l'analyse des données [33].

3 - Domotique :

La domotique ou maison connectée, c'est la manipulation de l'internet des objets dans une maison.

Les grandes entreprises telles que Nest, Netatom ont déployées des écosystèmes communiquant qui permettent de centraliser le contrôle de différents systèmes de votre maison.

Le principe de la domotique est de faire en sorte qu'une maison devienne intelligente indépendante et qu'elle réfléchisse par elle-même[33].

4 - La Santé :

Plusieurs applications dans l*e domaine de la santé* utilisent déjà l'internet des objets, machine a rayon X et imagerie, Porteuse Digital Health, Compteur D'énergie …etc.

Plus de 60% des hôpitaux mondiale utilisent déjà l’internet des objets, le secteur de la santé a connu pas mal d'applications permettant la liaison d'un patient à son docteur afin de récupérer des informations concernant sa maladie.

Les Objets connectées sont utilisés à la surveillance des établissement médicaux, les opérations chirurgicales, et les services de la géolocalisation[33].

5 - L’internet des objets dans le domaine de L’automobile :

Le marché des transports à déjà anticipé l’arrivée des objets connectés. Parmi les enjeux les plus fréquents que ce domaine fait naitre on retrouve la réduction des accidents et

des embouteillages, le partage entres voitures, le développement des offres de VTC et de TAX ou encore la gestion de flotte d'automobile. [33]..

6 - L'internet des objets dans le domaine de la sécurité :

Pour le cabinet en stratégie, ces entreprises vont rapidement se positionner comme des alliés des personnes dans leur domicile. En fournissant des données relatives à la consommation d'énergie des foyers, ces groupes vont apparaitre comme des arguments de factures contre les fournisseurs d'énergie où la précision sera difficile à tenir car ils seront probablement contraints d'accompagner leurs clients à une baisse énergétique des factures [33].

7 - L'internet des objets dans le domaine de l'industrie :

Le déploiement de L'IoT dans l'industrie sera certainement un grand support pour le développement de l'économie et le secteur des services, puisque L'IDO permettra d'assurer un suivi total des produits, de la chaine de production, jusqu'à la chaine de distribution en supervisant les conditions d'approvisionnements. Cette traçabilité de bout en bout facilitera la lutte contre la contrefaçon, la fraude et les crimes économiques transnationaux.

Certains éditeurs tels que SAP et CISCO montrant d'ores et déjà comment certaines zones industrielles comme le port d'Hambourg ont pu être équipés en puces et autres objets connectés. L'internet couvre un énorme nombre d'industries et utilise des cas qui s'étendent d'un seul dispositif contraint aux déploiements croisés de technologies intégrées de systèmes Cloud connectés en temps réel[33].

IV - Architecture de l'IoT :

L'Internet de Objets nécessite un modèle de référence qui permettrait de décrire la manière avec laquelle ces systèmes, ces réseaux et ces applications interagissent entre eux. En effet, un tel modèle aurait des avantages de :

- **Simplifier** :la compréhension de systèmes complexes découpés en parties plus compréhensibles
- **Clarifier** : en fournissant des informations supplémentaires et identifiant les niveaux de l'IoT en offrant une terminologie commune
- **Identifier** : où des types spécifiques de traitement sont optimisés dans les différentes parties du système

- **Standardiser** : pour créer les conditions d'une interopérabilité entre des produits IoT des différents fabriquants
- **Organiser** : rend l'IoT plus accessible et moins conceptuel.

Nous présentons maintenant les différentes architectures qui ont été proposées par certains chercheurs **:**

1 - Architectures à trois et cinq couches :

L'architecture la plus élémentaire est une architecture à trois couches [3–5], Elle a été introduite aux premiers stades de la recherche dans ce domaine. Il comporte trois couches, à savoir les couches perception, réseau et application.

- La couche de perception est la couche physique, qui possède des capteurs pour détecter et recueillir des informations sur l'environnement. Il détecte certains paramètres physiques ou identifie d'autres objets intelligents dans l'environnement.
- La couche réseau est responsable de la connexion à d'autres objets intelligents, périphériques réseau et serveurs. Ses fonctionnalités sont également utilisées pour transmettre et traiter les données des capteurs.
- La couche application est chargée de fournir des services d'applications spécifique à l'utilisateur. Il définit les diverses applications dans lesquelles l'Internet des objets peut être déployée, par exemple, les maisons intelligentes, les villes intelligentes et la santé intelligente.[36]

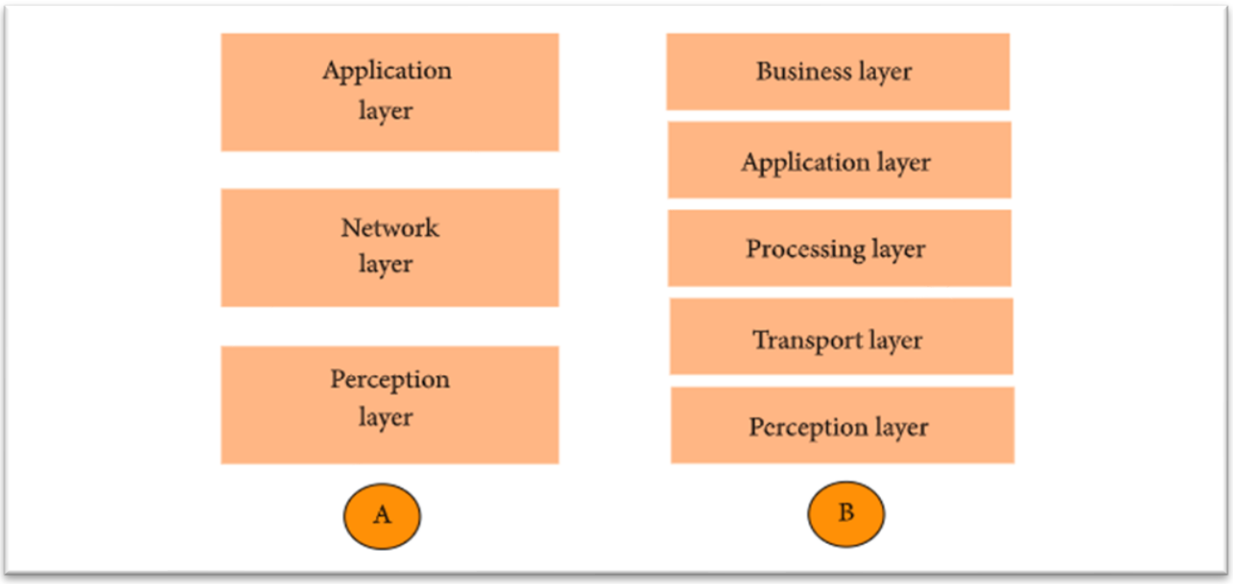


Figure 20 : Architecture de l'IoT (A : trois couches) (B : cinq couches) [82]

En ce qui concerne l'architecture à trois couches, elle définit l'idée principale de l'Internet des objets, mais elle n'est pas suffisante pour la recherche sur l'IoT car la recherche se concentre souvent sur des aspects plus fins de l'Internet des objets.

C'est pourquoi, nous avons beaucoup plus d'architectures en couches proposées dans la littérature. L'une est l'architecture à cinq couches, qui comprend en outre les couches de traitement et d'entreprise [3–6]. Les cinq couches sont les couches perception, transport, traitement, application et métier). Le rôle des couches de perception et d'application est le même que celui de l'architecture à trois couches.

- La couche de transport transfère les données du capteur de la couche de perception à la couche de traitement et vice versa via des réseaux tels que sans fil, 3G, LAN, Bluetooth, RFID et NFC.
- La couche de traitement est également connue sous le nom de couche *middleware*. Il stocke, analyse et traite d'énormes quantités de données provenant de la couche transport. Il peut gérer et fournir un ensemble diversifié de services aux couches inférieures. Il utilise de nombreuses technologies telles que les bases de données, le cloud comptine et les modules de traitement des mégadonnées.

- La couche métier gère l'ensemble du système IoT, y compris les applications, les modèles commerciaux et de profit et la confidentialité des utilisateurs. [36].

2 - Architectures basées sur le cloud et le brouillard (cloud and fog) :

Certaines architectures de systèmes, le traitement des données est effectué de manière centralisée à grande échelle par des ordinateurs en nuage. Une telle architecture centrée sur le cloud maintient le cloud au centre, Le cloud comptine bénéficie de la primauté car il offre une grande flexibilité et évolutivité. Il propose des services tels que l'infrastructure principale, la plate-forme, les logiciels et le stockage. Les développeurs peuvent fournir leurs outils de stockage, leurs outils logiciels, leurs outils d'exploration de données et d'apprentissage automatique ainsi que leurs outils de visualisation via le cloud.

Dernièrement, il y a une évolution vers une autre architecture de système, à savoir le calcul de brouillard, où les capteurs et les passerelles de réseau font une partie du traitement et de l'analyse des données. Une architecture de brouillard présente une approche en couches, comme le montre la figure 21, qui insère des couches de surveillance, de prétraitement, de stockage et de sécurité entre les couches physiques et de transport. La couche de surveillance surveille l'alimentation, les ressources, les réponses et les services. La couche de prétraitement effectue le filtrage, le traitement et l'analyse des données des capteurs. La couche de stockage temporaire fournit des fonctionnalités de stockage telles que la réplication, la distribution et le stockage des données. Enfin, la couche de sécurité effectue le chiffrement / déchiffrement et garantit l'intégrité et la confidentialité des données. La surveillance et le prétraitement se font en bordure du réseau avant d'envoyer des données vers le cloud[36].

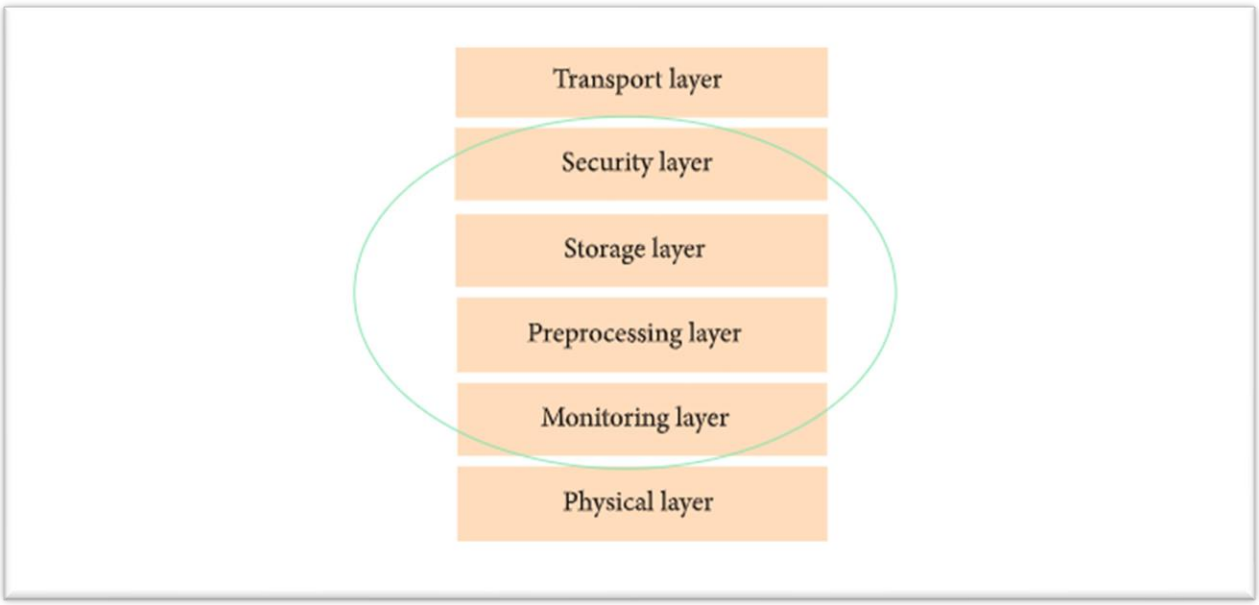


Figure 21 : Architecture de brouillard d'une passerelle IoT intelligente [37]

V - Les Protocoles de communication de l'internet Des Objets :

Un protocole de communication est responsable de :'' définir la politique et les règles et les procédures de communication des couche physique et la liaison du modèle OSI ,garce a ces protocoles on peut établir une connexion d'un objet à un réseau sans fil ou filaire qui permettre la transmission et la réception des données depuis l'internet à travers passerelle , Il existe de nombreuses options de passerelle, certaines aussi simples qu'un périphérique mobile (smart phone) co-localisé avec le point de terminaison IoT et communiquant via un RF protocole tel que Bluetooth-LE, ZigBee ou Wi-Fi.(Sennoun,) ''.

Quand on parle de la connexion d'un objet cela évoque les communications sans fils et les technologies telles que le WIFI, le Bluetooth, il existe pas mal de supports et dizaines de protocoles avec des caractéristique différentes (portée, débit ...etc.).

Avant que d'essayer d'adapter tous les protocoles IoT aux modèles d'architecture existants tels que le modèle OSI, ils ont divisé les protocoles en couches suivantes :

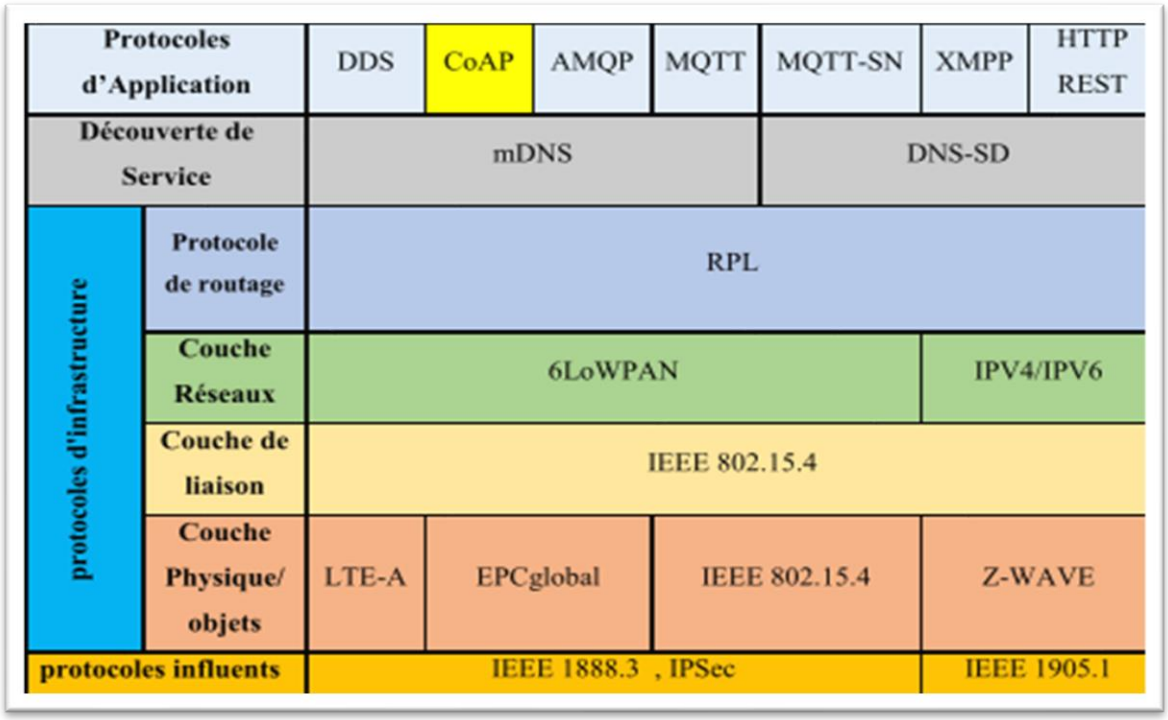


Figure 22 : Protocoles de communication de l'internet Des Objets[37]

VI - Protocoles d'infrastructure :

1 - Routing Protocol for Low Power and Lossy Networks (RPL) :

LIETF (Internet Engineering Task Force) à découvert l'importance de créer un nouveau groupe de travail pour trouver une solution de routage IPv6 pour les réseaux d'objets intelligents IP, le nouveau groupe appelé *ROLL (Routing Over Low power and Lossy).*

Le groupe de travail de routage IETF sur des liaisons à faible puissance et avec perte *(ROLL)* a normalisé un protocole de routage indépendant des liaisons basé sur IPv6 pour les nœuds à ressources limitées appelés RPL, RPL a été créé pour prendre en charge les exigences de routage minimales grâce à la création d'une topologie robuste sur les liaisons avec perte.

Ce protocole de routage est responsable de : '' prend en charge des modèles de trafic simples et complexes tels que multipoint à point, point à multipoint et point à point[34] ''.

1-1 - 6LoWPAN:

Pour pouvoir parler de protocole *6LoWPAN* nous devons savoir qu'est-ce qu'un WPAN ?

Les réseaux personnels sans fil de faible puissance *(WPAN)* sur lesquels de nombreuses communications IoT peuvent s'appuyer ont certaines caractéristiques spéciales différentes des anciennes technologies de couche liaison comme la taille limitée des paquets (par exemple, 127 octets maximum pour IEEE 802.15.4), diverses longueurs d'adresse et une faible bande passante, Il était donc nécessaire de créer une couche d'adaptation qui adapte les paquets IPv6 aux spécifications IEEE 802.15.4.

Ce protocole est développé par le groupe l'IETF comme norme en 2007. Le 6LoWPAN est la spécification des services de mappage requis par IPv6 sur des WPAN à faible puissance pour maintenir un réseau IPv6. [35]- [39]''.

1-2 - IEEE 802.15.4 :

Le protocole IEEE 802.15.4 a été créé pour spécifier une sous-couche pour le contrôle d'accès moyen (MAC) et une couche physique (PHY) pour les réseaux privés sans fil à faible débit (LR-WPAN) [40].

IEEE 802.15.4 a pour objectif de : '' prend en charge trois bandes de canaux de fréquence et utilise une méthode à spectre étalé en séquence directe (DSSS). Sur la base des canaux de fréquence utilisés, la couche physique transmet et reçoit des données sur trois débits de données : 250 kbps à 2,4 GHz, 40 kbps à 915 MHz et 20 kbps à 868 MHz. Des fréquences plus élevées et des bandes plus larges offrent un débit élevé et une faible latence tandis que les fréquences plus basses offrent une meilleure sensibilité et couvrent de plus grandes distances. Pour réduire les collisions potentielles, IEEE 802.15.4 MAC utilise le protocole CSMA / CA[33] ''.

1-3 - EPCglobal :

Plusieurs informaticiens ont pris le souci de clarifier ce terme afin de mieux le comprendre : '' Le code de produit électronique (EPC) est un numéro d'identification unique qui est stocké sur une étiquette RFID et est utilisé essentiellement dans la gestion de la chaîne d'approvisionnement pour identifier les articles. EPCglobal, en tant qu'organisation originale responsable du développement d'EPC, gère la technologie et les normes EPC et RFID. L'architecture sous-jacente utilise des technologies RFID basées sur Internet ainsi que des étiquettes et lecteurs RFID bon marché pour partager des informations sur les produits [41]. ''.

Cette architecture est : '' reconnue comme une technique prometteuse pour l'avenir de l'IoT en raison de son ouverture, de son évolutivité, de son interopérabilité et de sa fiabilité au-

delà de sa prise en charge des principales exigences de l'IoT telles que les ID d'objets et la découverte de services [42]''.

1-4 - LTE-A (Long Term Evolution—Advanced) :

Au niveau de la couche physique : '' le LTE-A utilise l'accès multiple par répartition en fréquence orthogonale (OFDMA) par lequel la bande passante du canal est partitionnée en bandes plus petites appelées blocs de ressources physiques (PRB).

Le LTE-A utilise également une technique à spectre étalé à porteuses multiples (CC) qui permet d'avoir jusqu'à cinq bandes de 20 MHz. L'architecture du réseau LTE-A repose sur deux parties essentielles. Le premier est le Coré Network (CN) qui contrôle les appareils mobiles et traite les flux des paquets IP.

L'autre partie est le réseau d'accès radio (RAN) qui gère les communications sans fil et l'accès radio et établit les protocoles du plan utilisateur et du plan de contrôle. Le RAN se compose principalement de stations de base (également appelées NodeB évoluées) qui sont connectées les unes aux autres par l'interface X2.

Le RAN et le CN sont connectés via l'interface S1. Les appareils mobiles ou MTC peuvent se connecter aux stations de base directement ou via la passerelle MTC (MTCG). Ils peuvent également avoir une communication directe avec d'autres appareils MTC[33]''.

1-5 - Z-Wave :

Z-Wave en tant que protocole de communication sans fil à faible puissance pour les réseaux domotiques (HAN) a été largement utilisé dans les applications de contrôle à distance dans les maisons intelligentes ainsi que dans les domaines commerciaux de petite taille .

Il a été développé par ZenSys (actuellement Sigma Designs) et a ensuite été utilisé et amélioré par Z-Wave Alliance[43].

Ce protocole permet de : '' couvre environ 30 mètres de communication point à point et est spécifié pour les applications qui nécessitent une transmission de données minuscule comme le contrôle de la lumière, le contrôle des appareils électroménagers, l'énergie intelligente et le CVC, le contrôle d'accès, le contrôle des soins de santé portable et la détection d'incendie.

Z-Wave fonctionne dans les bandes ISM (environ 900 MHz) et permet un taux de transmission de 40 kbps. Les versions récentes prennent également en charge jusqu'à 200 kbps.

Sa couche MAC bénéficie d'un mécanisme anti-collision. Une transmission fiable est possible dans ce protocole par des messages ACK optionnels. Dans son architecture se présentent des nœuds contrôleurs et esclaves.

Les contrôleurs gèrent les esclaves en leur envoyant des commandes. À des fins de routage, un contrôleur conserve une table de toute la topologie du réseau. Le routage dans ce protocole est effectué par la méthode de routage source dans laquelle un contrôleur soumet le chemin à l'intérieur d'un paquet[33]''.

VII - La sécurité et la protection de la vie privée (privacy)

Le niveau de l'acceptation des nouvelles technologies et des services offert par l'IOT au sein de la société est fortement lié au degré de fiabilité des informations et de protection des données privées des utilisateurs. Bien que plusieurs projets ont été lancés dans le but de trouver des solutions adéquates pour la protection de la privacy et d'assurer une protection rigoureuse aux utilisateurs finaux, à la confidentialité, la privacy et la gestion de la confiance [44]

VIII - Vulnérabilités et menaces dans l'internet des Objets

A cause de la forte intégration de l'IOT, les objets du quotidien deviennent des risques potentiels d'attaque sur la sécurité, l'ubiquité de L'IoT amplifiera les menaces classiques de la sécurité qui pèsent sur les données et les réseaux, de plus l'apparition de nouvelles menaces qui toucheront directement à l'intégrité des objets eux-mêmes, les infrastructures et processus et la privacy des personnes [44].

1 - Menaces sur les données et les réseaux

Le manque de la surveillance et de la protection physique des objets communicants peut engendres des attaques potentielles portées sur le matérielle telles que le vol, la corruption ou la contrefaçon de ces derniers pour la récupération des données qui sont stockées sur ces dispositifs ou pour interrompre le bon fonctionnement des réseaux ou des systèmes complexes qui les hébergent.

De plus, les transmissions sans fil sont réputées par leur forte vulnérabilité aux attaques de l'écoute passive et de déni de service. Les solutions cryptographiques existantes aujourd'hui ne sont pas adéquates pour tenir faces à ces problèmes cités à cause de la limitation de ressources des objets communicants, de ce fait, l'adaptation de ces dernières ou la conception de nouveaux modèles est une nécessité afin d'assurer les services de sécurité [44]

2 - Menaces sur la vie privée

De nombreux objets seront intégrés, portés ou même bien installés dans les lieux privés des personnes, ces objets présentent une potentielle menace pour la vie privée (privacy) de leurs utilisateurs. En effet, ces appareils électriques non seulement sont traçables, mais peuvent filmer, écouter ou même enregistrer leurs rythmes cardiaque ou respiratoire ainsi que la température du corps ou sa cinématique dans le but d'un malicieux [44].

3 - Menaces sur les systèmes et l'environnement physique des objets

Des objets malicieux connectés à un réseau ou intégrés dans un système complexe peuvent causer un dysfonctionnement quelconque, un déni de service ou autres types d'attaques à l'intégrité des données et les informations sensibles du système, ou pire encore prendre le contrôle du système en causant des importants dégâts [44].

IX - Attaques dans l'IoT

L'IoT est vulnérable à un nombre considérable d'attaques. Il existe diverses attaques sur des schémas d'authentification d'utilisateurs distants tels que le dictionnaire, men-in-the middle, le texte en clair, la carte à puce perdue, la modification, le déni de service (DOS), la divulgation de clé de session, l'emprunt d'identité, etc. Ces attaques peuvent être gênantes pour un utilisateur légitime lors de l'accès à un système dans un but spécifique. Une attaque de dictionnaire tente de deviner des mots de passe communs basés sur le dictionnaire. Une attaque men-in-the-middle est implémentée pour reconnaitre l'information. Une attaque en clair est utilisée lorsque le texte chiffré est volé. Une attaque perdue de carte à puce est introduite lorsqu'une carte à puce est perdue, puis un attaquant peut appliquer des procédures pour acquérir l'information. Une attaque de modification est implémenté pour modifier les informations ; en d'autres termes, l'attaquant modifie les informations puis retransmet les données à nouveau [45]. Ces attaques sont présentées dans le tableau 4

Nom de l'attaque	But et résultat de l'attaque	Menace	Active ou Passive
DoS	-Saturer un serveur ou bloquer le trafic. -rendre un service non disponible.	Intégrité. Disponibilité. Confidentialité.	Active
Man-in-the-Middle	- Intercepter les communications entre deux Parties contrôler la conversation. - ´écouter, modifier ou supprimer des données.	Intégrité. Confidentialité	Active

L'usurpation d'identité	- vol d'identité. - réaliser des actions frauduleuses. - prendre délibérément l'identité d'une autre personne Vivante.	Confidentialité Authentification.	Active
Footing	- épuiser la mémoire et l'énergie des nœuds - Saturer le réseau	Disponibilité.	Active
Les attaques de cartes à puce	- pouvoir accéder aux informations et aux secrets contenus dans la carte (code PIN, Clé(s) secrète(s) cryptographie(s), etc.…).	Physiques Logicielles	Active
Wardriving	- Utilisé pour pouvoir accéder à internet au nom D'une autre personne. - Parcourir tous les lieux où le Wifi est déployée afin De découvrir toutes les bornes Wifi existantes noter L'adresse géographique.	Confidentialité.	Passive
Sniffing	- Capturer les trames circulent local et afficher leur contenus (entêtes des paquets sur un réseau protocoles, id des user, MDP non crypté, etc.).	-confidentialité.	Passive

Tableau 4 : Type d'attaques dans l'IoT

X - Les attaques des routages RPL sur l'IoT :

1 - IETF RPL :

RPL est un protocole de couche réseau, assurant le routage sur la plupart des périphériques IoT. Dans le scénario le plus courant impliquant RPL, les nœuds du réseau sont connectés via des chemins d'accès à sauts multiples à un petit nombre de périphériques racine, qui sont responsables des tâches de collecte et de gestion de données.

RPL consiste en un ou plusieurs DODAGs (Destination Oriented Directed Acyclic Graphs) c'est-à-dire des graphes acycliques orientés dirigés vers une destination qui est la racine du réseau. Ces graphes sont dirigés de façon à éviter les boucles parce que chaque noeud dans le DODAG a un rang (distance de la racine), et ce rang doit diminuer en remontant dans le graphe vers la racine. RPL assure la QoS dans la couche réseau à partir d'une fonction objectif qui permet d'optimiser la topologie en fonction d'une contrainte/métrique comme la préservation de l'énergie, le chemin le plus court ou la qualité des liens.[46]

Pour construire un DODAG une racine envoie un message DIO (DODAG Information Object).

Une fois ce dernier reçu, chaque noeud RPL connaît l'ensemble de ses voisins et son rang suivant la fonction objectif Le rang d'un noeud correspond à son emplacement dans le graphe par rapport à la racine.

La valeur du rang augmente toujours en descendant dans le graphe, C'est donc la racine qui a le rang le plus petit dans le graphe.

Les DIOs sont envoyés à intervalles réguliers afin de maintenir les chemins de routage. Si un nœud reçoit des DIOs de voisins différents, c'est l'émetteur avec le meilleur rang (le plus petit donc) qui est choisi comme parent préféré.

C'est à dire le noeud vers lequel seront envoyés tous les messages à destination de la racine, et ainsi de suite dans le DODAG.

La fonction objectif doit être définie pour chaque implémentation du standard ***IETF ROLL***, car elle n'est pas fournie par le protocole, Chaque **DODAG** est identifié par un **ID**, le **DODAGID**, et le graphique est construit sur la base d'une métrique de rang. Cette métrique décroît de façon monotone le long du **DODAG** et vers le mote (nœud) cible.

Selon [47], **RPL** prend en charge trois modes de sécurité, à savoir:

- **Non sécurisé,** dans lequel les messages de base tels que DIS (Sollicitation d'informations DODAG), DIO (objet information DODAG), DAO (objet annonce DODAG) et DAOACK pour la configuration de maillage ne portent pas de sections de sécurité, reposant sur les protocoles de couche inférieure pour sécuriser les cadres.
- **Pré-installé,** dans lequel un nœud censé rejoindre le réseau possède une clé pré-installée, devenant soit un hôte ou un routeur, garantissant le message, la confidentialité, l'intégrité et l'authenticité.
- **Authentifié,** qui s'apparente au mode précédent, il est basé sur une pré-installation d'une clé, étant uniquement autorisé à devenir un hôte. Pour être promu routeur, il doit obtenir
- **une deuxième clé d'une autorité de clé,** ce qui peut authentifier que le demandeur est autorisé à être un routeur avant de lui fournir la deuxième clé.

Le réseau sécurisé utilise le mode pré-installé ou authentifié; de toute façon, il sera signalé avec un bit spécifique, le bit «A», sur la charge utile du paquet de couche 3 ou (couche

de réseau). Toutes les implémentations RPL doivent avoir l'algorithme CCM (Counter with Cipher Block Chaining-Message Authentication Code) CCM nécessite AES-128 comme algorithme cryptographique.

Lorsqu'un nœud a l'intention de rejoindre un réseau sécurisé, il est supposé qu'il a été configuré avec une clé partagée pour communiquer avec ses futurs voisins et la racine de RPL.

Pour ce faire, le nœud écoute les DIO sécurisées ou déclenche des DIO sécurisées en envoyant un DIS sécurisé.

Il existe des règles [47] qui définissent les valeurs spécifiques des messages DIS et DIO qui doivent être présentes sur chaque champ de chaque paquet.

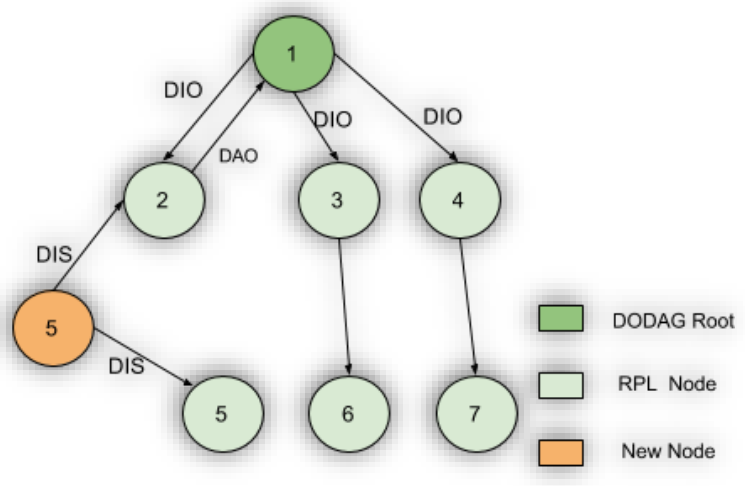


Figure 23:Exemple d'un réseau RPL (01 DODAG) [48]

2 - Attaques de RPL :

on proposons d'établir une taxonomie des attaques de routage contre le protocole RPL. Celle-ci prend en compte les objectifs de l'attaque et l'élément du réseau RPL qui est touché.

La taxonomie est présentée à la figure 24 et prend en compte trois catégories d'attaques de sécurité [48]:

2-1 - La première catégorie :

couvre les attaques visant l'épuisement des ressources du réseau (énergie, mémoire et puissance).

Ces attaques sont particulièrement dommageables pour ces réseaux contraints car elles réduisent considérablement la durée de vie des appareils et donc celle du réseau RPL.

2-2 - La deuxième catégorie :

comprend les attaques visant la topologie du réseau RPL[48], où elles **perturbent le fonctionnement normal du réseau** .

la topologie peut être sous-optimisée par rapport à une convergence normale du réseau ou un ensemble de nœuds RPL peut être isolé du réseau.

2-3 - La troisième catégorie :

correspond aux attaques contre le trafic du réseau, telles que les **attaques d'écoute** ou les **attaques de détournement**.[48]

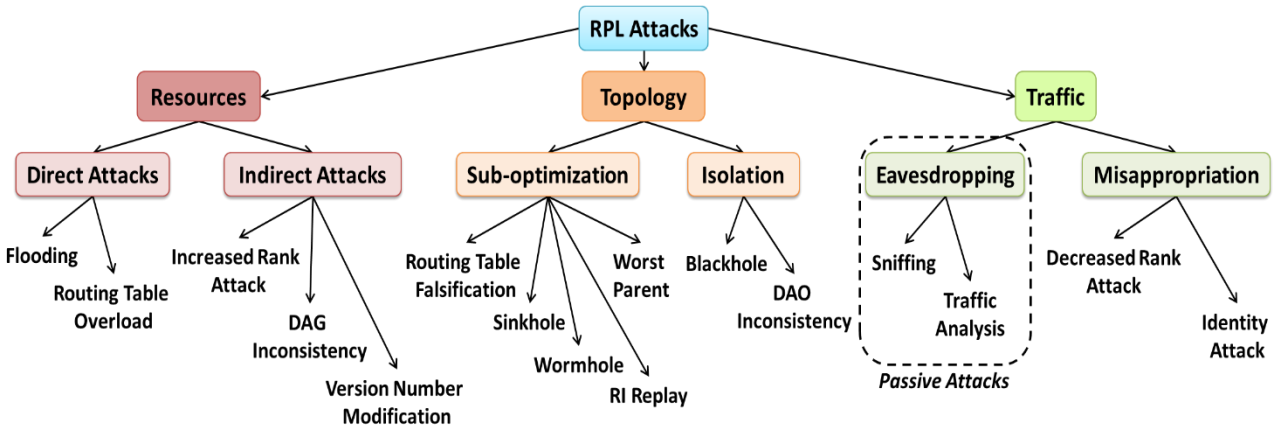


Figure 24:Taxonomie des attaques contre les réseaux RPL[48]

3 - Taxonomie :

Dans ce qui suit , nous entamons de façon détaillée ce qui a été mentionné précédemment [48]:

3-1 - La première catégorie :

concerne l'épuisement des ressources du réseau qui signifie que le but du nœud malveillant est de **surcharger la consommation d'énergie**, de **la mémoire** et/ou de **la puissance**.

Cela peut être fait en forçant les nœuds légitimes à effectuer des actions inutiles pour augmenter l'utilisation de leurs ressources ce qui peut avoir un impact sur la **disponibilité** du

réseau en encombrant les liens disponibles ou en rendant les nœuds incapables ce qui mènera à avoir une incidence sur la durée de vie du réseau .

Cette **catégorie** peut être subdivisée en deux sous-catégories [48]:

- **Des attaques directes :** dans lesquelles le nœud malveillant génère directement la surcharge perturbant le réseau.
- **Des attaques indirectes** :dans lesquelles le nœud malveillant provoque les autres nœuds pour leur faire générer de la surcharge.

3-2 - La seconde catégorie :

regroupe les attaques visant la topologie du réseau RPL,Le but de ces attaques est de **perturber** le **fonctionnement** normal du réseau, alors elles pourraient provoquer **l'isolement** d'un ou de plusieurs nœuds.

Cette catégorie peut également être subdivisée en deux sous-catégories[48]

- La **sous-optimisation** : qui signifie que le réseau convergera vers une forme non optimale, induisant de mauvaises performances.
- L'**isolation :** d'un nœud ou d'un sous-ensemble de nœuds, les coupant du reste du réseau, y compris du nœud racine.

3-3 - La troisième catégorie :

couvre les attaques contre le trafic du réseau.

Ces attaques visent à faire en sorte qu'un nœud malveillant s'introduise à l'intérieur du réseau, sans en perturber le fonctionnement ce qui entraîne une **fuite** d'informations en écoutant le trafic ou en se faisant passer pour des nœuds légitimes.

Cette catégorie se subdivise à nouveau en deux sous-catégories [48]:

- **l'écoute (passive)** :des informations qui sont transmises par le réseau
- **le détournement** :d'un nœud ou d'un ensemble de nœuds, notamment pour altérer les informations légitimes échangées.

4 - exemples d'attaques de routage :

Parmi les attaques de routage les plus importantes, on trouve les attaques de type "decreased rank" (DR), "hello-flood" (HF) et "version number modification" (VN), Black hole (BH)[48].

4-1 - Diminution du rang DR:

Consiste à faire de la publicité pour un rang inférieur afin que les nœuds légitimes se connectent au DODAG via l'attaquant ; cela peut bien sûr servir de base à des attaques de type : "trou noir" ou encore " les attaques d'écoute ".[49]

4-2 - Hello flood Attack HF :

Le protocole de routage nécessite que les nœuds voisins d'un même réseau échangent des messages HELLO pour indiquer leur présence et leur disponibilité, découvrir des routes et mettre à jour les tables de routage. Le nœud malveillant envoie un nombre énorme de paquets HELLO à différents nœuds pour se présenter comme voisin, afin qu'ils lui transmettent leurs données. [50]

4-3 - Version Number Attack VN:

Le but des attaques est d'augmenter le champ du numéro de version à l'intérieur des messages DIO et de les transmettre à ses voisins. En conséquence, une nouvelle construction DODAG est forcée de causer la perte de paquets de données, l'encombrement du réseau et l'épuisement des ressources des nœuds en raison de la surcharge du message de contrôle[51].

4-4 - Blackhole Attack BH:

Une attaque Blackhole dans un réseau signifierait qu'un ou plusieurs nœuds malveillants laisseraient tomber totalement ou partiellement les paquets de données qui y sont acheminés, ce qui entraînerait des perturbations dans le flux normal des données sur le réseau. Un nœud malveillant faussera les informations de routage, se présentera comme le meilleur chemin vers le nœud de contrôle (appelé node sink), pour forcer le passage des données par lui-même. Sa seule mission est alors de ne rien transférer, créant une sorte de puits ou de blackhole dans le réseau. L'intrus se place à un emplacement stratégique de routage dans le réseau et supprime tous les messages qu'il doit retransmettre, entraînant la suspension du service de routage du réseau dans les routes qui passent par le noeud du pirate. Si un nœud malveillant a la capacité d'usurper l'identité d'un nœud valide du réseau, il peut le faire lorsque le mécanisme de découverte d'itinéraire répond au nœud initiateur par un message de rediffusion routière en

annonçant un chemin avec un coût minimal au nœud demandé. Le nœud émetteur mettra alors à jour sa table de routage avec cette fausse route[52].

XI - La sécurité internet des objets

1 - La sécurité Technique

La cryptographie joue un rôle important, puisqu'elle permet d'adresser au moins partiellement les problématiques de confidentialité, d'intégrité, de disponibilité et de non-répudiation. Des outils cryptographiques appropriés et prêts à l'emploi existent aujourd'hui pour quasiment tous les besoins de sécurité imaginables. Plus globalement, il est vivement recommandé d'utiliser systématiquement les bonnes pratiques existantes à tous les niveaux, et d'éviter les mesures de sécurité '' originales '' ou '' créées sur mesure ''. Des composants réutilisables, fiables et éprouvés sont toujours bien plus sûrs qu'un code ou protocole créé spécialement pour la solution, et non éprouvé [53]

2 - Sur le physique de l'objet

Par exemple contre la copie physique ou celle de son micrologiciel.

- Le scellement du boîtier des objets connectés : il s'agit de '' verrouiller '' le boîtier de l'objet par (collage, thermocollage, soudure, ...) de façon à empêcher son ouverture normale, au détriment, il est vrai de pouvoir facilement réparer l'objet. Cela permet aussi de voir si l'intégrité physique de l'objet a été atteinte au premier coup d'œil.
- Le moulage des cartes et composants électroniques dans de la résine (si possible opaque aux rayons X) : cela empêchera l'identification des composants utilisés, ainsi que leur analyse par mesure ou débogage.
- La désactivation des ports de débogage et de lecture mémoire des composants pour empêcher l'analyse de leur comportement et des données traitées.
- L'utilisation de composants sécurisés pour le stockage des clés et les traitements cryptographiques : cela rend quasiment impossible l'extraction des secrets stockés dans l'objet.
- L'utilisation de clés et mots de passe tous différents dans chaque périphérique : pour éviter qu'une compromission sur un périphérique ne puisse compromettre l'intégralité du parc.
- Le chiffrement et l'obfuscation du micrologiciel : même s'il s'agit de sécurité par l'obscurité, jamais efficace à long terme, cela ralentit considérablement le travail

d'analyse et de rétro-ingénierie du micrologiciel, et donc son altération ou sa copie. [53]

3 - Sur les protocoles de communication

Il existe trois principales menaces récurrentes sur les communications : l'écoute passive, le brouillage (volontaire ou involontaire), et l'usurpation,:

Quelques règles pour qu'un un audit de la sécurité de la chaîne complète du flux de données et afin d'identifier l'ensemble des vulnérabilités pour bien comprendre les risques et les solutions à mettre en place par la suite :

- Le chiffrement des communications sensibles.
- L'authentification des objets les uns avec les autres : on parle alors d'authentification mutuelle, ce qui évite les attaques de type ‘‘ Man-in-the-Middle ‘‘. Habituellement, cette authentification est faite sous forme d'une procédure de pairage, ou par certificat électronique.
- L'utilisation de protocoles sécurisés d'échange de clés : le plus courant étant le protocole Diffue-Hellman, attention ce dernier nécessite auparavant d'avoir authentifié les périphériques concernés pour éviter l'interception par un tiers.
- L'utilisation de mécanismes anti-rejeu, comme les authentifications par challenge ou les numéros de séquence uniques et authentifiés (‘‘ nonce ‘‘ cryptographiques).
- L'utilisation de mesures anti-brouillage, comme l'étalement de spectre (‘‘ spread spectrum ‘‘) ou les sauts de fréquence (‘‘ frequency hopping ‘‘ / ‘‘ channel hopping ‘‘), utilisés dans certains protocoles comme Bluetooth.
- Certaines de ces mesures sont implémentées de base dans les protocoles sans fil. Celles qui ne sont pas implémentées peuvent souvent l'être au niveau de la couche applicative, donc restant à la charge du développeur. [53]

4 - Sur la sécurité applicative et système :

Enfin, la solution connectée peut-être attaquée par son maillon applicatif, que ce soit du côté client (smartphone, tablette, micrologiciel, …) ou du côté serveur (passerelle, serveur sur le cloud, …).

Voici des recommandations importantes à appliquées :

- Le choix et l'utilisation de briques applicatives répandues et régulièrement mises à jour.
- L'implémentation des contrôles côté serveur, car les contrôles côté client sont manipulables et contournables par un utilisateur malintentionné.
- Le filtrage et la validation de toutes les entrées, et si possible des sorties des traitements effectués. De manière générale, n'accorder aucune confiance aux données reçues.
- L'application d'une sécurité en profondeur, en donnant aux utilisateurs les privilèges minimums à leur utilisation de la solution et en redondant les mécanismes de sécurité eux-mêmes.
- L'utilisation de procédures et référentiels de durcissement système sur les serveurs de la solution : suppression des comptes et services non utilisés, applications de mots de passe forts sur tous les comptes, mises à jour régulièrement appliquées, monitoring actif des accès, veille permanente sur les vulnérabilités affectant les composants de la solution connectée [53].

XII - Objectifs de la sécurité :

La sécurité informatique d'une manière générale consiste à assurer que les ressources matérielles et logicielles d'une organisation sont uniquement dans le cadre prévu. Elle vise à assurer plusieurs objectifs, dont les cinq principaux sont : L'authentification, L'intégrité, la disponibilité et la non-répudiation [44]

Service	Solutions proposées	Remarques
Confidentialité	- VPN - TLS - DNS - Onion routing - PIR - Contrôle d'accès - Cloud computing	-La confidentialité est nécessaire pour protéger Les données sensibles. -Difficulté d'appliquer les solutions directement Au contexte de l'IoT en raison de l'extensibilité Et de contrôle d'accès. -Une politique de confidentialité doit être appliquée. -En général l'utilisateur privilégie du bénéfice du service face au risque de sa vie privée.
Authentification	- Mitiger les attaques par Déni de service. - Utilisation des techniques De détection d'intrusions	-Les limites de ressources dans l'IoT rendent difficile L'utilisation des algorithmes cryptographiques en Raison de leur consommation en termes de calcul Et de mémoire.

	Et d'authentification.	- Des recherches sont en cours afin de rendre ces algorithmes peu couteux et robustes à la fois.
Identification	- RFID - 6LoWpAN - IPv6 - Identification biométrique	- L'identification est cruciale pour l'IoT afin de faire correspondre les services avec leur demande. - L'identification permet de connaitre l'identité D'une entité. - Une identification robuste et scalable jouera Un rôle déterminant dans la sécurité de l'IoT. - Les ressources limitées de ces objets rendent aussi difficile l'implémentation de ces technologies.

Tableau 5 : Services de sécurité d'IoT avec Solutions proposées

1 - Transport Layer Security (TLS) :

basé sur une structure de confiance mondiale appropriée pourrait également améliorer la confidentialité et l'intégrité de l'IoT. Toutefois, comme chaque étape de délégation ONS nécessite une nouvelle connexion LS, la recherche d'informations serait affectée négativement par de nombreuses couches supplémentaires [54].

2 - Les extensions de sécurité DNS (DNSSEC) :

utilisent la cryptographie à clé publique pour signer des enregistrements de ressources afin de garantir l'authenticité de l'origine et l'intégrité´ des informations livrées. Toutefois, la DNSSEC ne peut assurer l'authentification de l'ONS que si l'ensemble de la communauté internet l'adopte [54].

3 - Onion Routing :

crypte et mélange le trafic Internet à partir de nombreuses sources différentes, c'est-`a-dire les données sont enveloppées dans plusieurs couches de cryptage, en utilisant les clés publiques des routeurs union sur le chemin de transmission. Ce processus empêcherait l'appariement d'un paquet de protocoles internet particulier à une source particulière. Cependant, le routage d'onion augmente les temps d'attente et entraîne ainsi des problèmes de performance [54].

4 - Les systèmes privés de récupération d'information (PIR) :

cachent quel client est intéressé par les informations, une fois que les EPCIS ont été localisés. Cependant, des problèmes d'évolutivité et de gestion des clés, ainsi que des problèmes de performance, se poseraient dans un système globalement accessible tel que l'ONS, ce qui rend cette méthode peu pratique [54].

5 - Peer-to-Peer (P2P) système :

est une autre méthode pour augmenter la sécurité et la confidentialité, qui montre généralement une bonne évolutivité et de performance dans les applications, P2P pouvait être basé sur des tables de hachage distribuées (DHT) [54].

6 - Contrôle d'accès :

assurer la confidentialité dans les systèmes de gestion de la connaissance. Une approche standard, qui correspond bien aux caractéristiques des environnements IoT, est représentée par le contrôle d'accès basé sur les rôles[55].

7 - Cloud comptant :

est une autre méthode pour assurer la confidentialité, si les périphériques IoT transmettent des données au cloud via la connexion HTTP, l'intégrité des données est atteinte. Dans le cas de HTTPS, la confidentialité et à l'intégrité sont abouties à la fois [56].

8 - IPv6 :

Une des technologies préconisées `a l'IETF pour l'interconnexion des réseaux de l'IoT est IPv6. Un des avantages majeurs est l'exploitation de l'immense capacité d'adressage de 128 bits d'IPv6 ce qui répondrait aux besoins d'adressage a très large échelle d'un IoT qui comporterait potentiellement plusieurs dizaines de milliards d'Objets [57].

9 - Identification biométrique :

R. Greenstadt et J. Beal ont proposé l'utilisation d'une imprégnation des Objets puis une identification biométrique continue pour la protection des Objets. Cette identification biométrique peut être diverse et variée comme les empreintes, l'image de la rétine, la fréquence de la voix, le mouvement, la reconnaissance du visage, etc. L'objectif est de permettre une

reconnaissance assez naturelle du propriétaire de l'Objet et ainsi éviter un tas de failles et d'attaques de sécurité par de tierces parties non l´légitimes à manipuler les Objets [58].

XIII - L'edge computing :

L'edge computing se définit comme une architecture informatique destinée aux environnements IoT, L'Edge computing est une méthode d'optimisation employée dans le cloud computing. Plutôt que de transférer les données générées par des appareils connectés IoT vers le Cloud ou un Data Center, il s'agit de traiter les données en périphérie du réseau directement où elles sont générées. Cette méthode réduit les besoins en bande passante des communications entre les capteurs et le centre de traitement des données en entreprenant les analyses et les connaissances au plus près de la source des données. En d'autres termes, les données sont traitées directement par le périphérique qui les génère (objet connecté, smartphone…) ou par un ordinateur / serveur local.

Jusqu'à récemment, le rôle de l'edge computing a principalement consisté à ingérer, stocker, filtrer et envoyer des données sur des systèmes dans le cloud. Toutefois, nous sommes aujourd'hui dans la situation où ces systèmes possèdent plus de puissance de calcul, de stockage et d'analyse pour consommer et agir sur les données dans le réseau machine.

1 - AVANTAGES ARCHITECTURES EDGE:

1-1 - Latence/Déterminisme:

Un traitement centralisé loin du cas d'utilisation crée des retards (dans la gestion du réseau et la vitesse de transmission) qui pourraient ne pas être acceptables pour certains traitements où une réaction en temps réel est nécessaire. [60]

1-2 - Données/bande passante:

Les données générées à la périphérie peuvent n'être utiles que quelques millisecondes après un événement ou n'avoir que peu de valeur (par exemple le flux vidéo d'une scène où rien ne se passe). Traiter les données à la source permet de les filtrer et de transmettre au datacenter uniquement les données ou métadonnées qui font sens, réduisant ainsi les besoins en bande passante. [60]

1-3 - Autonomie limitée:

Même dans les cas où la latence et la bande passante ne sont pas des exigences critiques, la nécessité de préserver un fonctionnement continu lorsqu'une connexion au système central est interrompue peut motiver un recours à l'Edge computing. [60]

1-4 - Confidentialité/Sécurité:

Les données captées à la périphérie deviendront toujours plus intimes (données de santé, reconnaissance faciale ou vocale, interactions dans des endroits privés) ou confidentielles (données critiques dans une usine). L'Edge sera utilisée pour répondre aux exigences réglementaires croissantes et à la protection de la vie privée par le traitement, le stockage et/ou l'élimination des données au plus près de la source. [60]

XIV - les défis de l'Internet des Objets (IOT)

Entre autres défis technologiques à relever, l'IoT a également des barrières à surmonter qui freinent aujourd'hui son développement[61].

1 - Alimentation des capteurs :

- Les capteurs devront être autosuffisants car leur nombre empêcherait tout entretien courant.
- Recherches en cours pour générer de l'énergie constamment (vibrations, lumière,).
- Il n'existe pas vraiment de standards techniques ce qui complique la gestion des équipements matériels pour des opérations de maintenance ou de réparation.[61]

2 - Identification :

- Chaque capteur devra avoir sa propre adresse IP.
- Le protocole IPv6 facilite la gestion des réseaux grâce à des fonctions de configuration automatiques et propose des fonctions de sécurité améliorées.
- De nombreux autres protocoles sont en cours de développement afin de pallier ce problème d'identification des objets.[61]

3 - Normes :

- Normes à établir dans les domaines de la sécurité, de la confidentialité, de l'architecture et des communications ainsi qu'un cadre juridique complet
- S'assurer que les paquets IPv6 peuvent être acheminés sur différents

types de réseau.

- Les normes de refonte de la 2G et le développement de la 5G pourraient toutefois bénéficier aux objets dits critiques**.**[61]

4 - Interopérabilité :

- Environ 40% de la valeur potentielle de l'IoT proviendra de la communication entre les différents systèmes de l'IoT et de l'intégration des données.
- Choisir des standards communs de communication ou des plateformes capables d'intégrer les données de plusieurs systèmes.
- Créer des algorithmes devront voir le jour afin de traiter la quantité phénoménale de données.[61]

Conclusion :

Parmi les défis majeurs de l'IOT c'est Le manque de normes dans l'Internet des Objets est très clairement un frein à la sécurité. Il manque tout d'abord des spécifications ouvertes sur beaucoup de protocoles sans fil et de systèmes embarqués existants. Mais au-delà, même lorsque des spécifications existent, il est rarissime de trouver des référentiels de sécurité sur ces technologies Pour résumer ce que nous venons de voir, nous devons protéger notre infrastructure contre ce type d'attaque, notre système de sécurité doit être robuste et fiable.

le chapitre suivant on va simuler certaines attaques de routage RPL et nous proposons une solution par l'utilisation de l'apprentissage automatique IDS.

Chapitre 4 : Contribution dans la détection des intrusions dans environnement IOT

Chapitre 04 : Contribution dans la détection des intrusions dans environnement IOT

Introduction :

Après avoir présenté le cadre théorique de notre travail, on se penche maintenant sur la deuxième partie dans le but de présenter notre simulation et extraire l'ensemble des données nécessaires pour notre test. On va construire un model IDS efficace aux attaques de routage dans l'environnement IoT

Ce chapitre est divisé en trois parties, dans la première nous présentons les travaux antérieurs et les différents outils utilisés, premièrement le simulateur Cooja pour simuler des attaques dans un environnement IoT afin d'extraire notre ensemble des données, et enfin nous présentons notre modèle et les différents résultats d'expérimentation avec notre proposition inclus dans ce travail par l'utilisation de Weka.

I - Travaux connexes :

Il existe plusieurs études sur l'attaque des protocoles de routage pour l'IdO en utilisant IDS :

1. Dans les recherches de : Abdelkader Alem, Youcef Dahmani,, Bendaoud Mebarek au titre : Skyline Computation for Improving Naïve Bayesian Classifier in Intrusion Detection System :

   - Ce document propose un IDS hybride à deux couches basées sur l'opérateur Skyline et le classificateur bayésien naïf.
   - le classificateur le plus approprié a été identifié par le calcul Skyline sur la base de trois critères, à savoir la précision, le taux de détection et le taux de fausses alertes , ensuite, les résultats ont été intégrés par le classificateur bayésien naïf dans la décision finale.
   - Pour vérifier son efficacité, l'IDS proposé a été testé sur le le célèbre ensemble de données KDD.
   - La structure principale du modèle proposé est composée de deux niveaux :
   - le premier comprend les meilleurs classificateurs basé sur les trois dimensions (précision, taux de détection et taux de fausses alarmes).

- Le deuxième niveau contient un classificateur bayésien naïf qui intègre les sorties du premier niveau pour prendre la décision finale.
- Le choix des meilleurs classificateurs du premier niveau basé sur un calcul Skyline (opérateur Skyline) sur trois critères principaux : maximiser la précision, maximiser le taux de détection, et minimiser le taux de fausses alertes.

Les résultats montrent que cette proposition se traduit par une meilleure performance si elle donne le meilleur taux de détection pour les attaques rares, elle préserve également un taux de détection et une précision élevés.

Cela est vrai même si il est comparé à des travaux bien connus dans la littérature qui utilisent exactement le même ensemble de données d'apprentissage et de test.[62]

2. Dans les recherches de Mridula Sharma, Haytham Elmiligi, Fayez Gebali et Abhishek Verma au titre : **Simulating Attacks for RPL and Generating Multi-class Dataset for Supervised Maching** :
    - leur document se concentre sur l'analyse des menaces à la sécurité en matière de RPL et sur les attaques qui pourraient affecter le réseau du CPS (Cyber Physical System)[1].
    - Ils présentent un nouveau Framework pour simuler les attaques RPL en utilisant le contiki-Cooja.et ils avaient simulés quatre attaques différentes à l'aide de ce Framework.
    - Pour la mises en œuvre de l'expérimentation , ils choisissent d'utiliser quatre attaques différentes : l'attaque "hello flood", l'attaque "DODAG Information Solicitation" (DIS), l'attaque "increased version" et l'attaque "reduced rank".

[1] Système cyber-physique est une francisation directe de l'anglais "cyber-physical system" (CPS) qui correspond à un système intégrant de l'électronique et du logiciel, des capteurs et des actionneurs et dotés de capacité de communication. Un système cyber-physique est autonome et "embarque" de ce fait les éléments mentionnés précédemment. Un CPS interagit avec son environnement dans lequel il prend des données, les traite et au travers d'une boucle de rétroaction contrôle ou influence le processus auquel il est associé. Les CPS sont utilisés pour contrôler et piloter les processus physiques et "augmentent" ainsi ces processus de fonctionnalités nouvelles. De par ses capacités de communication, un CPS peut agir en collaboration avec d'autres systèmes et/ou échanger des données avec des systèmes distants. La communication peut être filaire via un bus industriel par exemple et/ou sans fil. Lorsqu'un CPS utilise les technologies de communication de l'internet, il devient une brique de base de l'internet des objets.

- Ils analysent les caractéristiques extraites des paquets de trafic du réseau et propose un nouveau modèle d'apprentissage machine. En utilisant plusieurs techniques de réduction des caractéristiques, le nombre de caractéristiques requises pour la classification des attaques est réduit de 58 à 21, soit une réduction de 63,7% à économiser l'énergie de traitement et de communication.

- L'ensemble de caractéristiques choisi montre une efficacité accrue dans la détection de diverses attaques à l'aide de trois classificateurs différents, à savoir Naive Bayes, RandomForest et le et C4.5 .

Leurs résultats expérimentaux montrent que ils pouvaient atteindre une précision de classification de 99,33% en utilisant le classificateur Random-forest. [63]

3. Dans les recherches de : Abd Mlak Said, Aymen Yahyaoui, Faicel Yaakoubi, et Takoua Abdellatif au titre : **Learning Based Rank Attack Detection for Smart Hospital Infrastructure**

- Dans cet article, ils proposent un système de détection d'intrusion "IDS" pour les hôpitaux intelligents

- Ils proposent un système de détection des attaques RPL basé sur les anomalies contre un réseau IdO et specialement le RPL en utilisant des machines à vecteurs de support.

- Les hôpitaux auxquels ils intéressent sont confrontés à de nombreux défis tels que la résilience des services, l'interopérabilité des actifs et la protection des informations sensibles.

- Ils exécutent quatre scénarios de simulation :

  - **Scénario 1 :** réseau IdO sans mote (nœud) malveillant.
  - **Scénario 2 :** réseau IdO avec 1 mote malveillant placé au hasard.
  - **Scénario 3:** réseau IdO avec deux motes malveillants placés de manière aléatoire.
  - **Scénario 4 :** réseau IdO avec 4 motes malveillants placés de manière aléatoire.

- L'IDS choisi est centralisé et basé sur les anomalies en utilisant un algorithme d'apprentissage automatique SVM.

- Pour évaluer la précision de l'IDS proposé, ils se basent sur la consommation d'énergie comme paramètre, et ils recueillent des données de suivi de la puissance par mote en

termes d'énergie radio, d'énergie radio de transmission, d'énergie radio RX de réception et d'énergie radio INT interférée.

Les résultats obtenus montrent que le l'efficacité de l'approche par une précision de détection sera plus élevée et plus précise lorsque le nombre de nœuds malveillants augmente.[64]

4. Dans les recherches de : Yavuz, F. Y., Devrim, & Ensar, G. Ü. L. (2018) au titre : **Deep learning for detection of routing attacks in the internet of things. International Journal of Computational Intelligence Systems, 12(1), 39-58.**

- Cette étude est une preuve de concept vers l'application d'apprentissage approfondi pour la sécurité de l'IdO
- Ils proposent une méthode de détection des attaques de routage pour l'IdO basée sur l'apprentissage approfondi.
- Le principal problème dans ce domaine est le manque d'ensembles de données et la qualité des données disponibles. Nos ensembles de données sur les attaques sont produites par simulation, en utilisant le code d'un capteur réel et l'implémentation du protocole RPL de Contiki-RPL.
- Dans leur étude, ils utilisent le simulateur Cooja-IdO, afin de générer des données d'attaque hautement précises dans des réseaux IdO dont la taille allant de 10 à 1000 nœuds.
- Ils proposent une méthodologie de détection d'attaques nettement évolutive et basée sur l'apprentissage approfondi pour la détection des attaques de routage IdO qui sont des attaques de catégorie restreinte, de type "hello-flood" et de modification de numéro de version, avec une grande précision et exactitude.
- Ils notent que l'application de l'apprentissage approfondi à la cybersécurité dans l'IdO nécessite la disponibilité de données consistantes sur les attaques IdO
- Ils ont construit en outre, un réseau neuronal profond des modèles formés à l'aide des ensembles de données de l'IRAD avec les informations d'évaluation : l'exactitude, la précision et les taux de rappel.

Ils parviennent enfin, à obtenir jusqu'à 99%, sur la base des Scores F1 et le score du test AUC.[65]

5. Dans la recherche de Yukai Yao, Yang Liu, Yongqing Yu, Hong Xu, Weiming Lv, Zhao Li, Xiaoyun Chen au titre de **Un algorithme SVM effectif basé sur le clustering de K-means :**

   - Dans ce document, ils se consacrent à résoudre le problème de réduire la taille de l'ensemble de l'apprentissage avec la méthode du regroupement.
   - Ils utilisent l'approche du regroupement K-means pour sélectionner les quelques échantillons les plus instructifs, qui sont utilisés dans la construction de leur ensemble d'entraînement effectif
   - Ils utilisent La machine à vecteurs (SVM) , cet algorithme de classification populaires , le SVM se consacre à trouver l'hyperplan de séparation entre deux classes, ce qui peut donner une capacité de généralisation remarquable pour ce type d'activité. Afin de trouver le l'hyperplan optimal, ils prennent la plupart des enregistrements comme étant leur data set. Cependant, la séparation l'hyperplan n'est déterminé que par quelques échantillons cruciaux (Support Vectors, SVs),
   - Leur algorithme travaille pour sélectionner les échantillons informatifs utilisant l'algorithme de classification des K-means et le classificateur SVM est construit grâce à l'apprentissage sur ces échantillons sélectionnés.

Les résultats des expériences montrent que leur algorithme atteint l'objectif de réduire l'échelle de l'ensemble de données, et réduit considérablement le temps d'apprentissage et de prédiction, tout en assurant la capacité de généralisation de l'algorithme K-SVM.[66]

Notre contribution :

Nous avons généré un ensemble de données par des simulations équivalentes à la vie réelle en utilisant le célèbre simulateur open source Contiki/Cooja, en raison de la non-disponibilité des ensembles de données publiques sur les attaques IdO.

I - La Simulation :

1 - Le simulateur Cooja Contiki :

Contiki est un système d'exploitation flexible et léger pour les réseaux de capteurs, open source, écrit en C et peut être utilisé dans des systèmes commerciaux et non commerciaux. Il fonctionne avec un minuscule microcontrôleur à faible coût et développe des

applications qui utilisent efficacement le matériel et qui fournissent une communication sans fil standardisée à faible consommation pour la variété des plates-formes de matériel .

Contiki dispose l'un des outils majeurs appelé Cooja qui permet aux développeurs de tester leur code avant de s'exécuter sur le matériel cible. Cooja est un simulateur logiciel conçu pour les réseaux de capteurs sans fil, il est open source, construit en java, capable d'exécuter des programmes C, C+++, supporte IPV4, IPV6, ainsi que les derniers standards pour les réseaux sans fil basse consommation tels que 6LoWPAN, RPL et permet le déploiement de nombreux types de moteurs comme Z1, Skymote, MicaZ etc.(*Cooja Simulator*,)

Nous expliquons maintenant les étapes suivis pour accéder au simulateur :

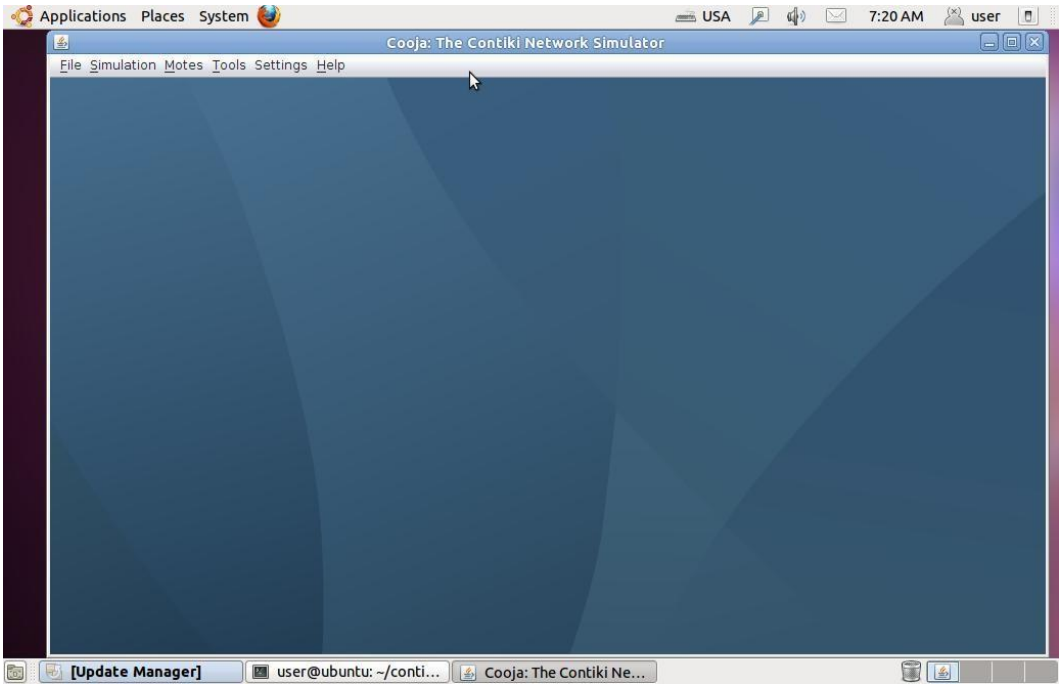


Figure 25 : Premier affichage Cooja

On clique sur File (Fichier), ensuite sur New Simulation (Nouvelle simulation) et l'écran illustré à la Figure 26 s'affiche à nouveau. Il n'est pas nécessaire de modifier les paramètres de cet écran.

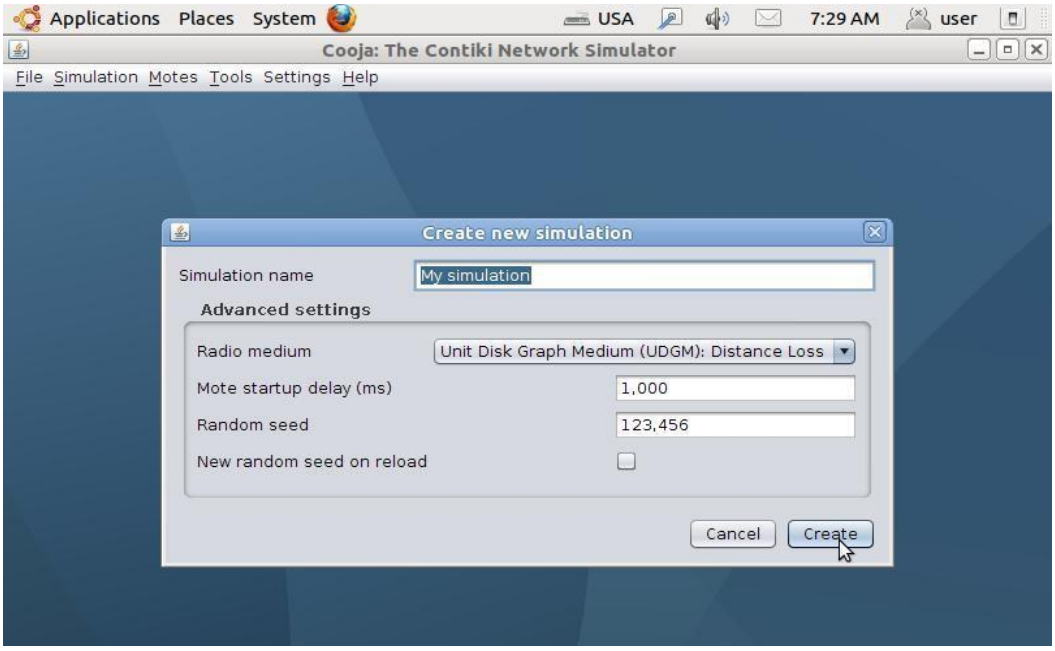


Figure 26 : Création d'une nouvelle simulation

Le bouton Create permet de lancer l'écran de simulation initial, comme le montre la Figure 27.

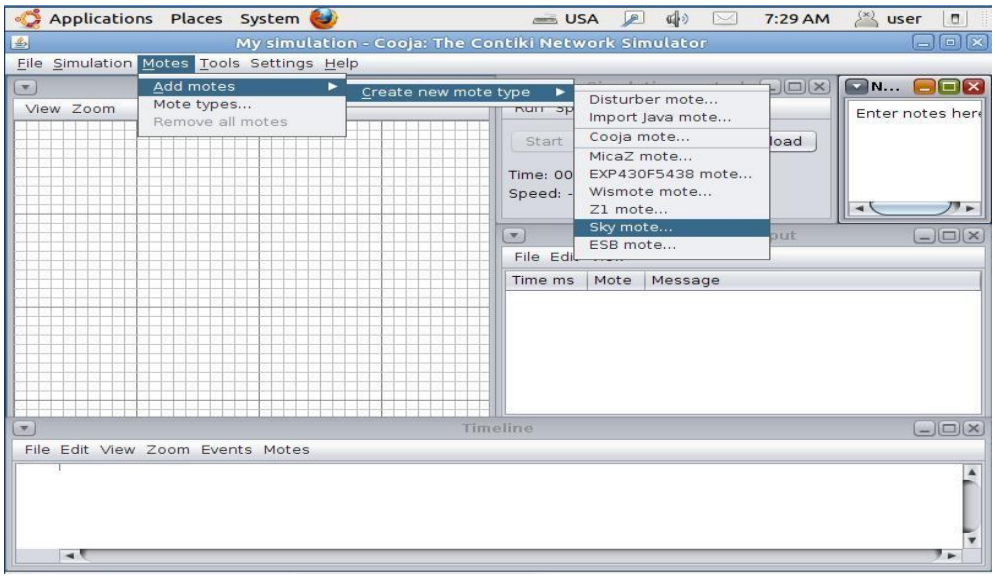


Figure 27 : Écran initial de simulation Cooja

Pour le moment, rien à faire et ceci est dû à l'absence des motes dans le réseau. Celles-ci sont ajoutées en cliquant sur Motes, on cliques sur Add motes, (Ajouter des motes) pour créer un nouveau type de Mote et ensuite on tapes Sky Mote à partir du menu qui en résulte, comme le montre la Figure 28.

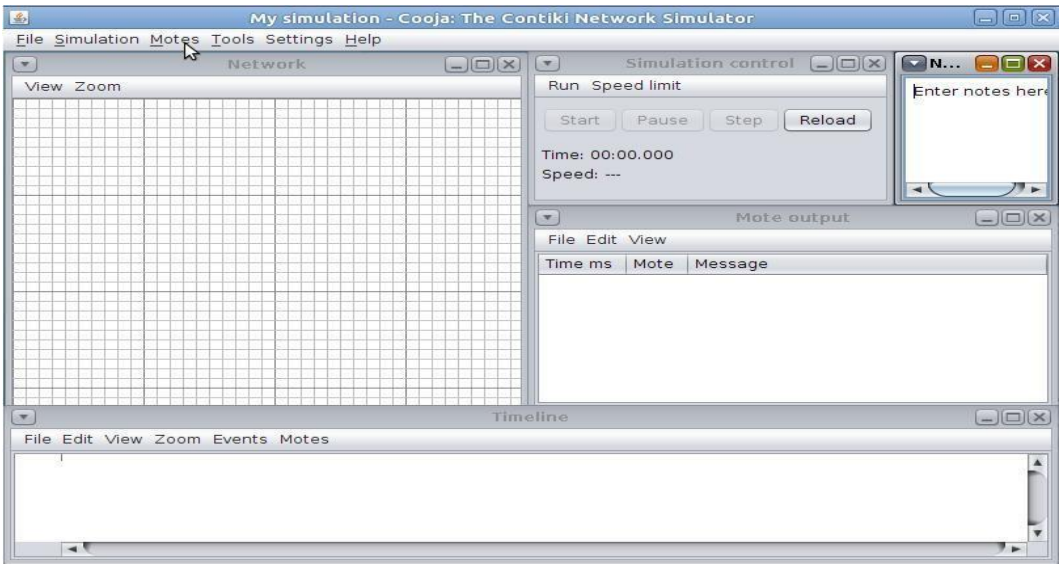


Figure 28 : Ajouter Motes

La Sky Mote est la plus simple forme des Motes à utiliser dans un WSN et l'idéal pour les configurations initiales dans une simulation Cooja. L'écran qui en résulte est affiché à la Figure 29.

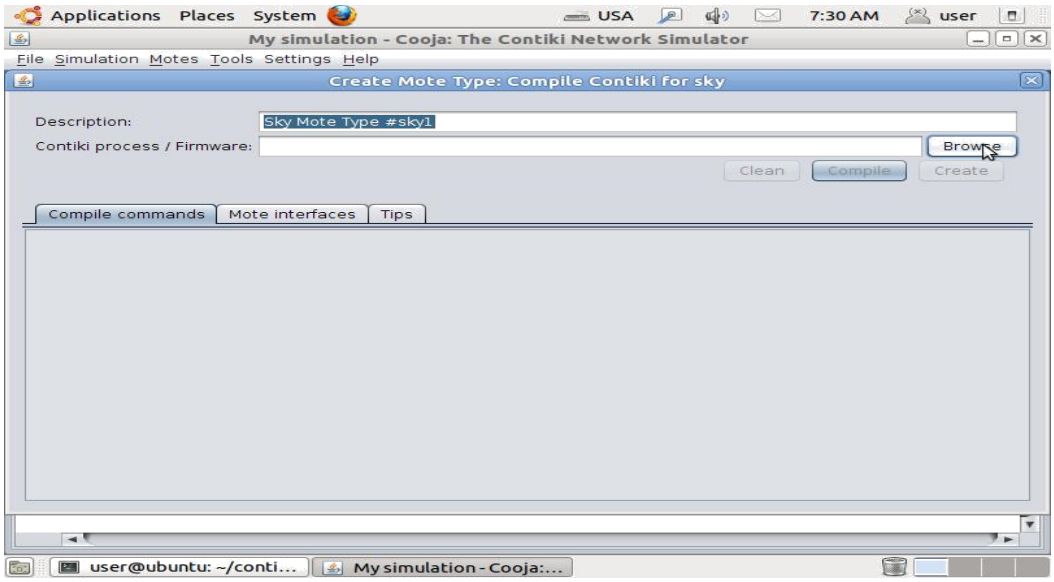


Figure 29 : parcourir le Mote

Comme on peut le distinguer dans la Figure 30, il y a un dossier exemples et c'est là que se trouve le Firmware, avec de très nombreuses options disponibles. Comme cet exemple qui implique l'utilisation de RPL, par le chemin sélectionné :

/home/user/contiki/examples/ipv6/rpl-collect/udp-sink.c. udp-sink.c. udp-sink c'est le firmware en langage C du mote qui va maintenant être créé.

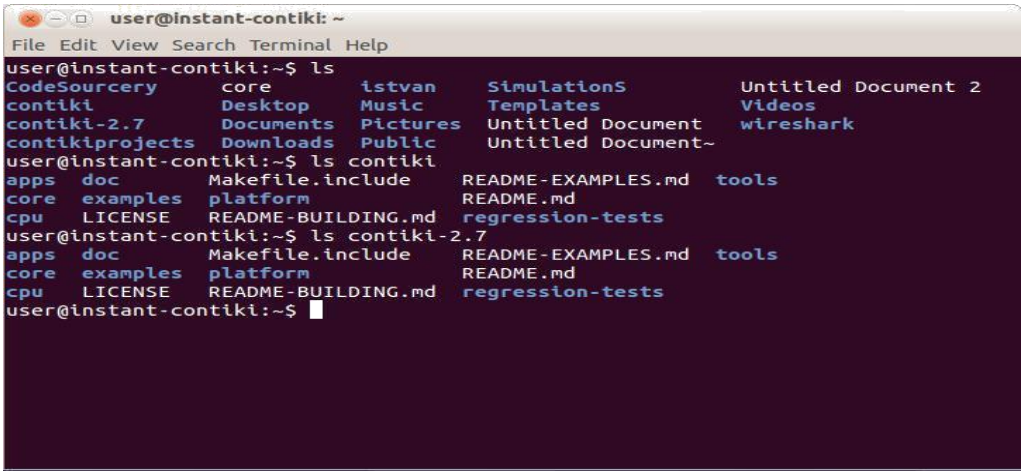


Figure 30 : Les fichiers contiki

On cliques sur Clean (Nettoyer) pour effacer toute compilation précédente de la mote, puis sur Compile. Il en résultera une sortie comme est illustré dans la Figure 31 qui montre l'issue de compilation. Il y aura toujours un code d'avertissement en rouge, à condition qu'il n'y a pas des erreurs en rouge, à la fin de la sortie le mote est compilé avec succès.

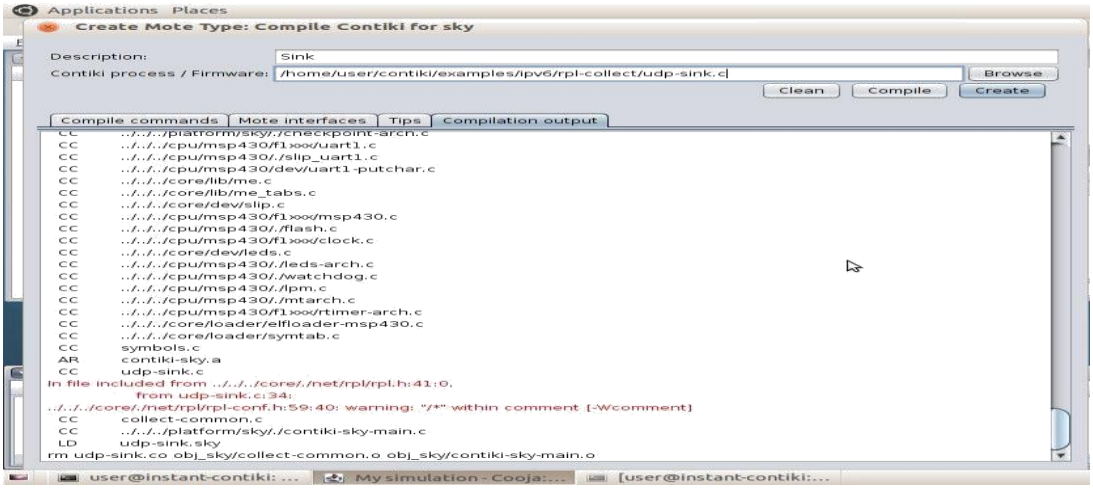


Figure 31 : Compilation de Mote Cooja

On clique maintenant sur Create pour faire apparaître l'option permettant de créer le nombre de Motes requises. Une case apparaîtra comme sur la figure 32.

Comme il s'agit d'une Mote Sink qui est nécessaire, cliquez sur Ajouter des Motes et une Mote est ajoutée. Ce processus doit être répété pour que les Motes expéditeur s'ajoutent avec le chemin du firmware Mote :

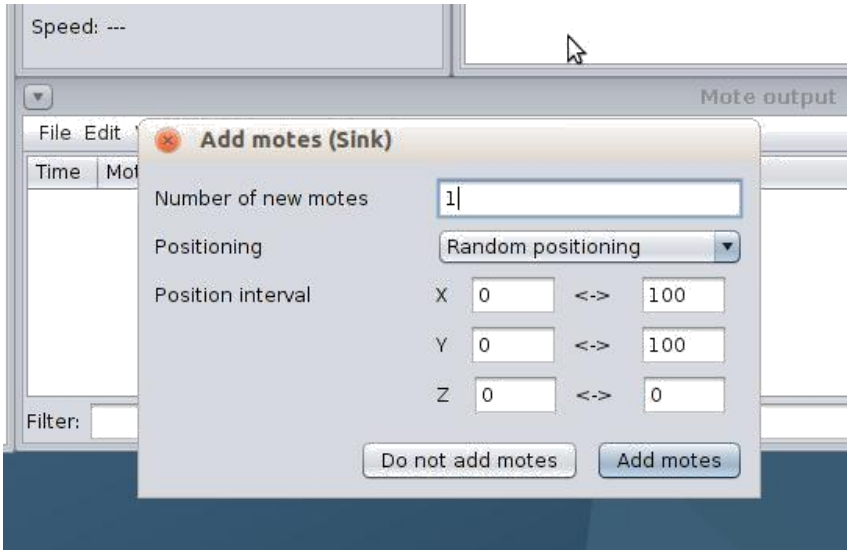


Figure 32 : Ajouter des Motes Cooja

Une fois qu'on clique sur le bouton Ajouter des Motes, un écran similaire à celui de la Figure 33 est affiché, montrant les Motes du réseau avec le numéro 1 étant Sink Mote.

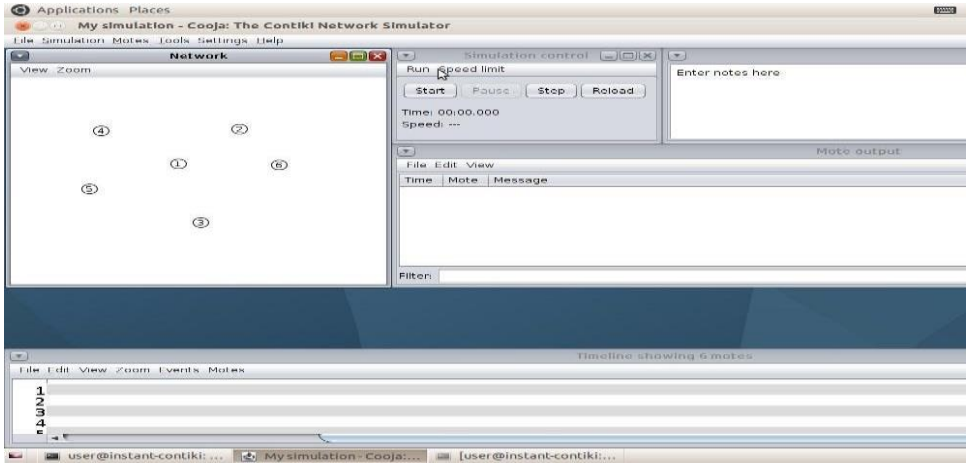


Figure 33 : Topologie initiale crée

2 - Simulation des attaques :

nous avons simulé les attaques de routage les plus importantes, "decreased rank" (DR), "hello-flood" (HF) et "version number modification" (VN), Black hole (BH)[48].

2-1 - Simulation d'attaque hello-flood :

L'objectif de cette partie est de démontrer que ce type d'attaque peut avoir un impact dramatique sur un réseau WSN par un épuisement énergétique important.

Le nœud malveillant commence immédiatement à envoyer des messages DIS à ses voisins, puis déclenche la réinitialisation des messages DIO et des minuteries de maintien.

2-2 - Configuration :

Le réseau WSN contient :

- 1 nœud racine de type root-dummy construit sur un Z1
- 10 capteurs de type capteur-mannequin construits sur un Z1
- 1 grain malveillant de type malware-sensor construit sur un Z1

Les capteurs sont répartis sur une zone de 200,0 mètres de côté et centrés autour du nœud racine à une distance minimale de 20,0 mètres et une distance maximale de 200,0 mètres. Ils ont une portée de transmission maximale de 50,0 mètres et une portée d'interférence maximale de 100,0 mètres.

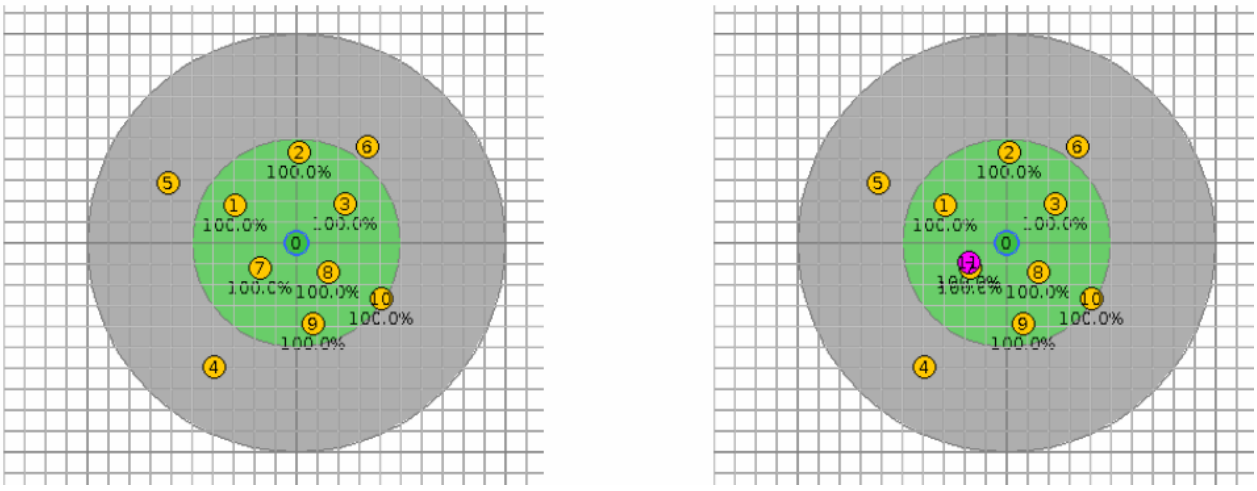


Figure 34 : Configuration WSN sans le malveillant et avec le malveillant

2-3 - Capture de paquets de données:

La première étape à réaliser est donc l'obtention de données représentatives du problème à résoudre. Dans notre cas c'est le trafic du réseau WSN. Nous avons capturé de façon continue le trafic de tout réseau à l'aide d'un outil "radio messages" proposé par Contiki.

Les données circulant dans le réseau sont stockées dans un fichier pcap que nous avons utilisé pour construire l'ensemble de données, comme le montre la figure 35.

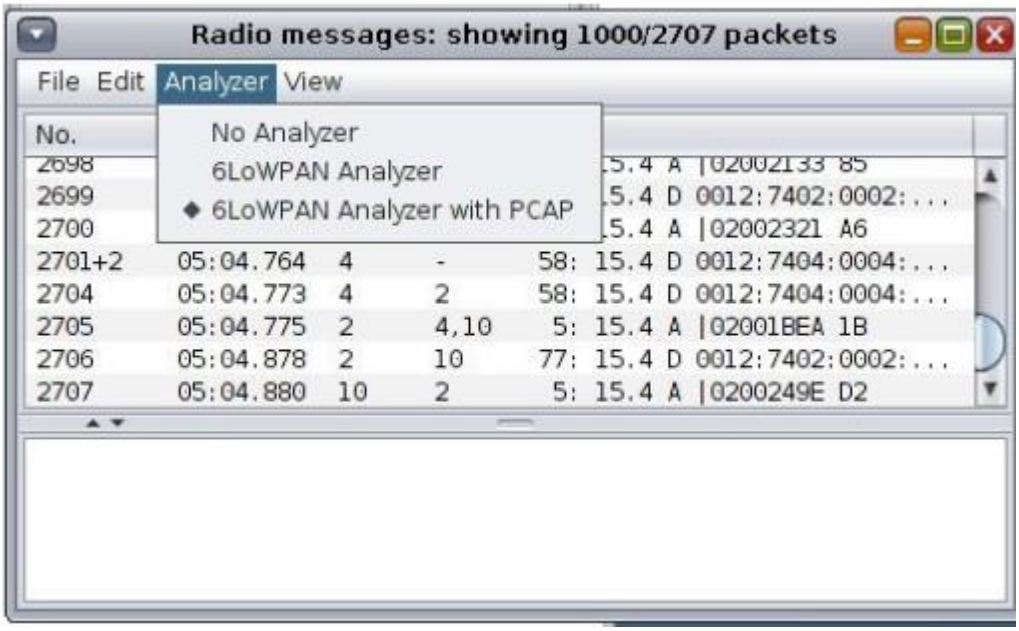


Figure 35:Capture de paquets de données sur cooja

II - Apprentissage automatique :

L'apprentissage automatique (« *machine Learning* ») est une méthode utilisée en intelligence artificielle. Il s'agit des algorithmes (*procédures traduites en langages informatiques*) qui analysent un ensemble de données afin de déduire des règles (une étape dite « entraînement ») constituant des connaissances permettant d'analyser de nouvelles situations. De grands ensembles de données (des *big data*) sont nécessaires pour l'entraînement des algorithmes d'intelligence artificielle.

III - Les types du ML :

L'apprentissage automatique vous permet d'entraîner les ordinateurs à agir de manière indépendante afin que nous n'ayons pas à rédiger des instructions détaillées pour l'exécution de certaines tâches. Pour cette raison, l'apprentissage automatique apporte une grande valeur pour n'importe quel domaine, mais tout d'abord, il fonctionnera bien là où il y a la science des données. L'apprentissage du Machine Learning est basé sur ces trois principaux types :

1 - Apprentissage Supervisé :

Dans ce type d'apprentissage, nous disposons d'un ensemble de données contenant des caractéristiques, mais chaque exemple est également associé à une étiquette (Label) ou à une cible.

Parmi les modèles phares de ce type d'apprentissage nous comptons les Réseaux de neurones, cette technique a été utilisée avec succès dans la reconnaissance de formes, Traitement automatique de la langue et d'autres applications innovantes .

l'apprentissage supervisé en deux sous catégories :la classification et la régression. [68]

2 - Apprentissage Non-Supervisé :

L'apprentissage non supervisé consiste à apprendre à un algorithme d'intelligence artificielle (IA) des informations qui ne sont ni classées, ni étiquetées, et de permettre à cet algorithme de réagir à ces informations sans supervision. [69]

3 - L'apprentissage par renforcement :

(RL pour Reinforcement Learning) fait référence à une classe de problèmes d'*apprentissage automatique*, dont le but est d'apprendre, à partir d'expériences successives, ce qu'il convient de faire de façon à trouver la meilleure solution. [69]

4 - Taxonomie par l'usage ou l'objectif :

Parmis les tâches d'apprentissage automatique les plus répandues nous retrouvons:

- Trouver la valeur d'une propriété d'un phénomène qui dépend des valeurs d'autres propriétés de ce phénomène (ou d'autres instances du phénomène). Il s'agit de la régression.
- Rechercher des frontières pour séparer les éléments de données en groupes. Lorsque les groupes sont imposés dès le départ (ils font partie des entrées de l'algorithme), la tâche est appelée classification. Lorsque les groupes doivent être déterminés par l'algorithme, nous appelons la tâche partitionnement (ou clustering).
- Optimiser les processus de décision.
- Découvrir des règles d'association (apprentissage de règles d'association).

- Désassembler une mesure (souvent une série temporelle) en ses composantes basiques. C'est la séparation aveugle de sources.

- Réduire le nombre de variables d'entrée d'un problème d'apprentissage. Cette tâche est appelée réduction de la dimensionnalité.

5 - Les Algorithmes de classification :

5-1 - SVM (Support Victor Machine) :

SVM est un algorithme d'apprentissage automatique supervisé qui peut être utilisé à la fois pour des problèmes de classification ou de régression. Cependant, il est principalement utilisé dans les problèmes de classification. Dans cet algorithme, nous plaçons chaque donnée sous forme de point dans un espace à n dimensions (où n est le nombre de caractéristiques que vous avez …), la valeur de chaque caractéristique étant la valeur d'une coordonnée particulière. Ensuite, nous effectuons la classification en trouvant l'hyper-plan qui différencie très bien les deux classes[70]

5-2 - Arbre j48 :

J48 est une méthode à base d'arbre de décision, l'objectif de ce type de méthode est de construire une fonction de classement représentable par un arbre qui est construit en partant de la racine et en allant vers les feuilles. On cherche à discriminer les exemples selon leur classe et en fonction des attributs considérés comme des meilleurs parmi les autres au sens d'un critère donné [71].

Une méthode très efficace d'apprentissage supervisé. Partitionne un ensemble de données en des groupes les plus homogènes possible du point de vue de la variable à prédire. On prend en entrée un ensemble de données classées, On fournit en sortie un arbre où : chaque nœud final (feuille) représente une décision (une classe) chaque nœud non final (interne) représente un test. Les branches représentent les résultats des tests Chaque feuille représente la décision d'appartenance à une classe des données vérifiant tous les tests du chemin menant de la racine à cette feuille.

5-3 - Random Forest :

Forêt aléatoire crée plusieurs arbres de décision et les fusionne pour obtenir une prédiction plus précise et plus stable. Comme je l'ai déjà mentionné, Random Forest est un ensemble d'arbres de décision, mais il existe quelques différences. Si vous entrez un jeu de données

d'apprentissage avec des entités et des étiquettes dans un arbre de décision, il formulera un ensemble de règles, qui seront utilisées pour effectuer les prédictions. Par exemple, si vous souhaitez prédire si une personne cliquera sur une publicité en ligne, vous pouvez collecter la publicité de la personne sur laquelle vous avez cliqué dans le passé et certaines fonctionnalités décrivant sa décision. Si vous mettez les caractéristiques et les étiquettes dans un arbre de décision, des règles seront générées. Ensuite, vous pouvez prédire si la publicité sera cliquée ou non. En comparaison, l'algorithme Random Forest sélectionne de manière aléatoire des observations et des entités pour créer plusieurs arbres de décision, puis effectue la moyenne des résultats [71].

5-4 - Les réseaux bayésiens naïfs :

Les réseaux bayésiens sont des outils de représentation de connaissances en présence d'incertitude. Le succès de ces modèles est fortement lié à leur capacité de représenter et de manipuler des relations de (in)dépendance qui sont importantes pour une gestion efficace des informations incertaines. Les réseaux bayésiens utilisent une représentation basée sur le conditionnement, où les connaissances sont structurées sous la forme d'un graphe acyclique orienté. Les nœuds représentent des variables et les arcs qui codent le lien causal (ou l'influence) entre ces variables. L'incertitude est représentée au niveau de chaque nœud en explicitant toutes les probabilités conditionnelles attachées aux valeurs associées à ce nœud sachant celles de ses parents. Cette incertitude exprime la force de la relation de causalité entre les variables. Une simple variante des réseaux bayésiens est appelée réseaux bayésiens naïfs. Ces réseaux ont une structure unique qui se compose de deux niveaux seulement. Le premier contient un seul nœud parent qui n'est pas observé et le second plusieurs enfants de ce nœud correspondant aux nœuds observés [72].

Réseaux bayésiens naïfs travaillent sous la forte hypothèse d'indépendance entre les nœuds enfants dans le contexte de leur parent. L'utilisation des réseaux bayésiens naïfs est assurée en considérant le nœud parent comme un nœud caché précisant à quelle classe appartient chaque objet de la base de données et les nœuds enfants représentent les différents attributs spécifiant cet objet. En présence d'un ensemble d'apprentissage on doit juste calculer les probabilités conditionnelles puisque la structure du graphe est unique.

Ce calcul peut être résumé comme suit :

- Les probabilités conditionnelles pour les attributs discrets sont calculées à partir des fréquences en comptant combien de fois chaque valeur d'attribut apparaît avec chaque valeur possible du nœud parent.

$$P(c_i / A) = \frac{f(A/c_i)}{f(c_i)} \quad (1)$$

- Une fois le réseau quantifié, il peut être utilisé pour classer de nouveaux objets étant donné leurs valeurs d'attributs en utilisant la règle de Bayes exprimée par :

$$P(c_i/A) = \frac{P(A_1/c_i) * P(c_i)}{P(A)} \quad (2)$$

- Où ci est une valeur possible de la classe C et A est l'évidence totale sur les attributs. L'évidence A peut être vue comme un vecteur d'instances a1, a2, ans relatifs aux attributs a1, a2, …, an respectivement. Puisque les réseaux bayésiens naïfs travaillent sous l'hypothèse que ces attributs sont indépendants (Sachant le nœud parent C), leur probabilité jointe peut être calculée comme suit [72]:

$$P(c_i/A) = \frac{P(a_1/c_i)\ P(a_2/c_i)\ \ldots.P(a_n/c_i) * P(c_i)}{P(A)} \quad (3)$$

6 - Validation et mesure performance :

6-1 - Matrice de confusion :

En Machine Learning, l'évaluation de la qualité de la classification est faite avec différentes mesures comme les faux positifs et les taux négatifs, la matrice de confusion, la précision, rappel et F-Mesure .

La matrice de confusion est une technique d'évaluation appliquée à tout type de problème de classification. Elle affiche les quatre valeurs (vrai positif, vrai négatif, faux positif et faux négatif) d'une manière dont la relation entre elles est facilement compréhensible comme le montre le tableau 6:

		Prédiction de la classe	
		Classe négative (normal)	**Classe positive (attaque)**
Classe actuelle	Classe négative (normal)	Vrai négative (VN)	Faux positif (FP)
	Classe positive (attaque)	Faux négative (FN)	Vrai positif (VP)

Tableau 6 :Matrice de confusion

6-2 - Métrique d'évaluation :

Les métriques de performance d'un IDS comprennent le taux d'exactitude, le taux de fausse alerte et le taux de détection des attaques, elles sont définies comme suit :

- Taux d'Exactitude : montre à qu'elle point le système est exacte, c'est le nombre des cas bien classés sur le nombre de type de tous les cas.

$$\boldsymbol{Exactitude} = \frac{VP + VN}{VP + VN + FP + FN}$$

- Taux de détection : mesure le taux des attaques détectées par un IDS dans un environnement donné et pendant une durée donnée. C'est le nombre des attaques détecté sur le nombre des attaques existants dans le corpus.

$$\boldsymbol{DR} = \frac{VP}{VP + FN}$$

- Les fausses alarmes : ce critère mesure le taux de fausses alertes générées par un IDS dans un environnement donné et pendant une durée donnée. C'est le nombre des alertes généré comme attaque sur le nombre des types classés comme normal existants dans le corpus.

$$\boldsymbol{FAR} = \frac{FP}{VN + FP}$$

- la précision : c'est-à-dire la proportion de prédictions correctes parmi les points que l'on a prédits positifs. C'est la capacité de notre modèle pour qu'il ne déclenche d'alarme que pour un vrai incendie.

$$Précision = \frac{\mathbf{VP}}{\mathbf{VP} + FP}$$

Pour évaluer un compromis entre rappel et précision, on peut calculer la "F-mesure", qui est leur moyenne harmonique.

- Rappel =DR.

$$F - mesure = 2 \times \frac{Précision \times Rappel}{Précision + Rappel} = \frac{2\mathbf{VP}}{2\mathbf{VP} + FP + FN}$$

7 - Génération de notre ensemble de données :

Dans ce chapitre nous allons d'abord expliquer les outils utilisés dans la simulation des différents scénarios de la communication en réseau IOT, l'environnement de la simulation est le célèbre Cooja , l'outil de simulation multicouche (application, système d'exploitation et couche de code machine) intégré dans le d'exploitation Contiki qui est en effet un flexible pour capteurs miniatures en réseau.

Les capteurs du réseau simulé fonctionnent avec le système d'exploitation Contiki et implémentent le protocole RPL , le Contiki permet de charger et de décharger des programmes et des services individuels sur les capteurs simulés.

C'est pour cette raison que nous avons effectué une simulation de chaque attaque comme mentionné ci-dessus, en exécutant le code réel des capteurs dans le simulateur, en utilisant une machine virtuelle avec 06 Go de RAM et 2 V-CPU monter sur un PC de Processor: Intel(R) Core(TM) i7-4500U CPU , Fréquence de base est 1.80GHz (4 CPUs), et La fréquence Turbo maxi est 2.4GHz.

Le Contiki comprend l'environnement d'exécution Java 64 bits en plus du système d'exploitation Ubuntu 64 bits.

Nous avons construit de différentes topologies de réseau pour simuler les attaques de routage de l'IOT, ensuite nous avons simulé ces scénarios par le biais du simulateur de réseaux Cooja, en évitant de produire l'ensemble de données synthétiques, vu que Cooja permet d'exécuter du code RPL réel sur les nœuds simulés, et permettre de prendre les messages radio transmis dans le réseau simulé sous forme d'un fichier PCAP que nous devons le convertir ensuite en format CSV (Comma Separated Values) afin de le traiter en mode texte par le biais de notre bibliothèque de prétraitement de données Python, et Après, un processus d'extraction de fonctionnalités sera appliqué sur les fichiers CSV générés.

Après avoir simulé les scénarios, des ensembles de données ont été produits sous forme de fichiers PCAP. Nous avons disséqué le fichier PCAP en CSV en utilisant Wireshark et en l'introduisant dans le prétraitement. Un échantillon de l'ensemble de données brutes est présenté dans le tableau 7.

No.	Time	Source	Destination	Protocol	Length	Info
228	4,932379	fe80::c30c:0:0:b	fe80::c30c:0:0:0	ICMPv6	76	RPL Control (Destination Advertisement Object)
204	2,334325	fe80::c30c:0:0:0	ff02::1a	ICMPv6	97	RPL Control (DODAG Information Object)
194	1,001214	fe80::c30c:0:0:1	ff02::1a	ICMPv6	64	RPL Control (DODAG Information Solicitation)
21626	87,740032	fe80::c30c:0:0:4	fe80::c30c:0:0:a	ICMPv6	102	RPL Control (DODAG Information Object)

Tableau 7 :Un échantillon de l'ensemble de données brutes capturés

7-1 - Les attaques et Les fonctionnalités de données (d'attributs) pour la phase de l'apprentissage automatique :

RPL (Routing Protocol for Low-Power and Lossy Net-works) est un protocole de routage IPv6 arborescent pour 6LoWPAN. Il crée des graphiques acycliques dirigés orientés vers la destination (DODAG), appelés arbre DODAG.

Chaque réseau a un ou plusieurs nœuds racine DODAG comme nœud central , et chaque réseau a un identifiant unique DODAG ID à identifier. En outre, chaque nœud a un numéro de

rang et une table de routage en raison des numéros de rang des autres nœuds. Le numéro de rang est utilisé pour déterminer la distance entre le nœud et la racine.

Dans le protocole RPL, il existe trois types de paquets de contrôle : DODAG Information Object (DIO), Destination Advertisement Object (DAO) et DODAG Information Solicitation (DIS). Les paquets DIO sont d'abord envoyés par le nœud de base (ou racine) sous forme de paquets de diffusion pour établir l'arbre DODAG. Les autres nœuds reçoivent les paquets DIO et créent leur table de routage en sélectionnant leur nœud parent. Ils envoient les paquets DAO au nœud parent, en demandant la permission de se connecter au nœud parent.

Le nœud parent accepte cette offre en renvoyant un paquet DIO ACK. Un nouveau nœud envoie des paquets DIS pour rejoindre l'arbre DODAG. Si un nouveau nœud rejoint l'arbre, tous les nœuds envoient à ce nouveau un paquet DIO pour réformer DODAG (ou topologie du réseau).

Les attaques de routage ont lieu au niveau de la couche réseau.

Parmi les attaques de routage les plus importantes, on trouve les attaques de type "decreased rank" (DR), "hello-flood" (HF) et "version number modification" (VN), Black hole (BH) (pour plus de détails voire chapitre 03) .

7-2 - Prétraitement des données et extraction de fonctionnalités :

Pour les attaques BH, HF et VN,DR nous avons généré de différents scénarios d'attaque.

Les simulations nous ont permis de produire des ensembles de données brutes, Les détails des scénarios (chaque fois change la position et les adresses IP des nœuds) et ces ensembles de données mentionnés précédemment sont énumérés dans le tableau 8.

Scénario	Nœud malveillant	Nœud normal	Nombre de paquets
Hello flood	1	10	182 029
	4		191 412
	4		106 472
	4		201 029
	4		90 853
Version Number	1	10	169 406
	4		161 182
	4		88 041
	4		142 107
	4		142 177
Blackhole	1	10	126 781
	4		161 852
	4		67 258
	4		65 934
	4		69 388
Decreased Rank	1	10	94763
	3		113937
	3		135409
	3		11051
Normal	0	10	131 380
			101 037
			100 125
			149 616
Total			2 803 239

Tableau 8: détails des scénarios de simulation

Après , avoir obtenus des fichiers de données brutes de la simulation. Cependant, les fichiers de données ne seront pas suffisants pour être inclus dans l'algorithme d'apprentissage automatique , car l'ensemble de ces données brutes comprend des informations telles que l'adresse des nœuds source/destination et la longueur des paquets, ce qui provoque du bruit et de l'Overfitting (le sur-apprentissage) dans l'algorithme d'apprentissage automatique. Pour cette raison, nous avons implémentés un prétraitement des données et un algorithme d'extraction de fonctionnalités en utilisant le langage Python et ces bibliothèques Pandas33 , ces dernières effectuent les opérations de prétraitement qui sont nécessaires pour faciliter l'extraction de fonctionnalités .

Nous avons mis en place une structure de dictionnaire pour traiter un grand nombre de nœuds, tout on choisissant de ne pas calculer de statistiques globales sur le temps total simulé ou le nombre total de paquets, parce que ce type de calcul pourrait nuire au calcul des poids des fonctionnalités extraites.

Nous avons divisé toute la simulation en périodes de temps, ou fenêtres d'une durée de 1000 ms.

Avant ce processus, il est nécessaire de trier les ensembles de données par temps de simulation, car une séquence correcte du temps de simulation des paquets est nécessaire pour calculer correctement la valeur des fonctionnalités .

Le pseudocode de l'algorithme de prétraitement des données et d'extraction des fonctionnalités est fourni dans l'algorithme 1 (figure 36).

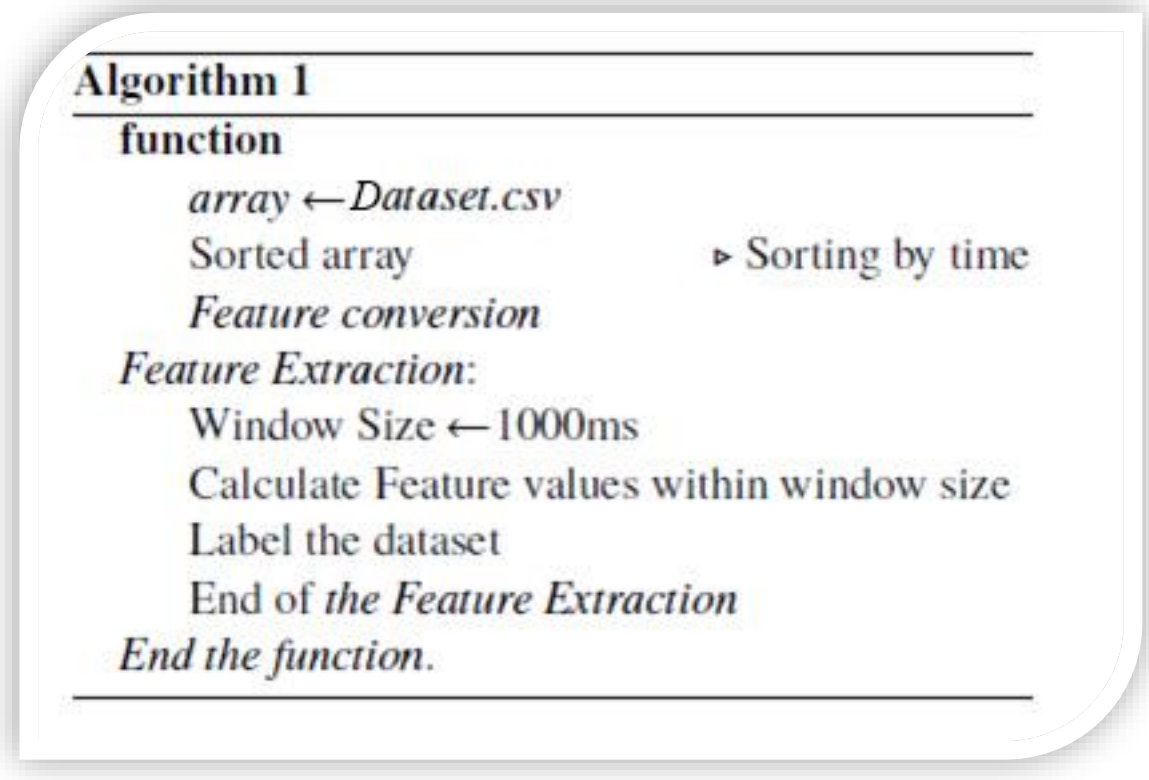


Figure 36:l'algorithme de prétraitement des données

Les ensembles de données brutes comprennent des types de données qui ne peuvent pas être traités par l'algorithme d'apprentissage automatique, comme les adresses IP. Les adresses de source et de destination sont converties du format IPv6 en ID de nœud. Par exemple : l'**adresse IPv6** 2001:0db8:3c4d:0015:0000:d234::3eee:0011 peut être raccourcie en 11.

Les paquets de diffusion sont traités dans un ensemble de données brutes, si l'adresse de destination est ff02::1a, cela signifie que le nœud source envoie des paquets de diffusion. Cette

valeur est convertie en 99 pour éviter toute coïncidence avec un autre nœud nous avons également encodé les protocoles respectivement en 1, 2, 3.

Le DAO est utilisé dans le protocole RPL pour envoyer des informations de destination unicast sur les parents sélectionnés. DIO est le type de message le plus important dans le protocole RPL. Il conserve le rang actuel du nœud, détermine la meilleure route à travers le nœud de base en utilisant des métriques spécifiques comme la distance ou le nombre de sauts. Un autre type de message est DIS. Les nœuds utilisent DIS pour recevoir les messages DIO.

L'extraction de fonctionnalités a permis de produire un total de 13 attributs .

Ces valeurs sont calculées comme suit. Tout d'abord, nous calculons le nombre de paquets transmis et reçus pour chaque nœud en 1000 ms dans un délai donné. Ensuite, nous divisons ces valeurs à 1000 ms et obtenons le taux de transmission et le taux de réception pour chaque nœud, respectivement ,pour toutes les périodes.

On calcule le temps de transmission et de réception de chaque paquet. Le temps total de la durée de chaque paquet de transmission et de réception en 1000 ms. Ensuite, on calcule le temps moyen de transmission et de réception pour chaque nœud, Le nombre de paquets de contrôle transmis de chaque nœud (concernent les paquets de contrôle : DAO, DIO et DIS) est calculé dans la taille de fenêtrage, 1000 Ms.

7-3 - Nouvelles fonctionnalités intégrées :

Nous nous sommes focalisés sur le paramètre de la consommation d’énergie dans le but d’évaluer la précision de l’IDS que nous avons proposé. Nous présentons en graphique les informations recueillies lors du suivi de la puissance de chaque mote, en termes d’énergie (radio ON), énergie d’émission (radio TX : mode émission), énergie de réception (radio RX : mode réception) et enfin l’énergie INT (radio interférée) :

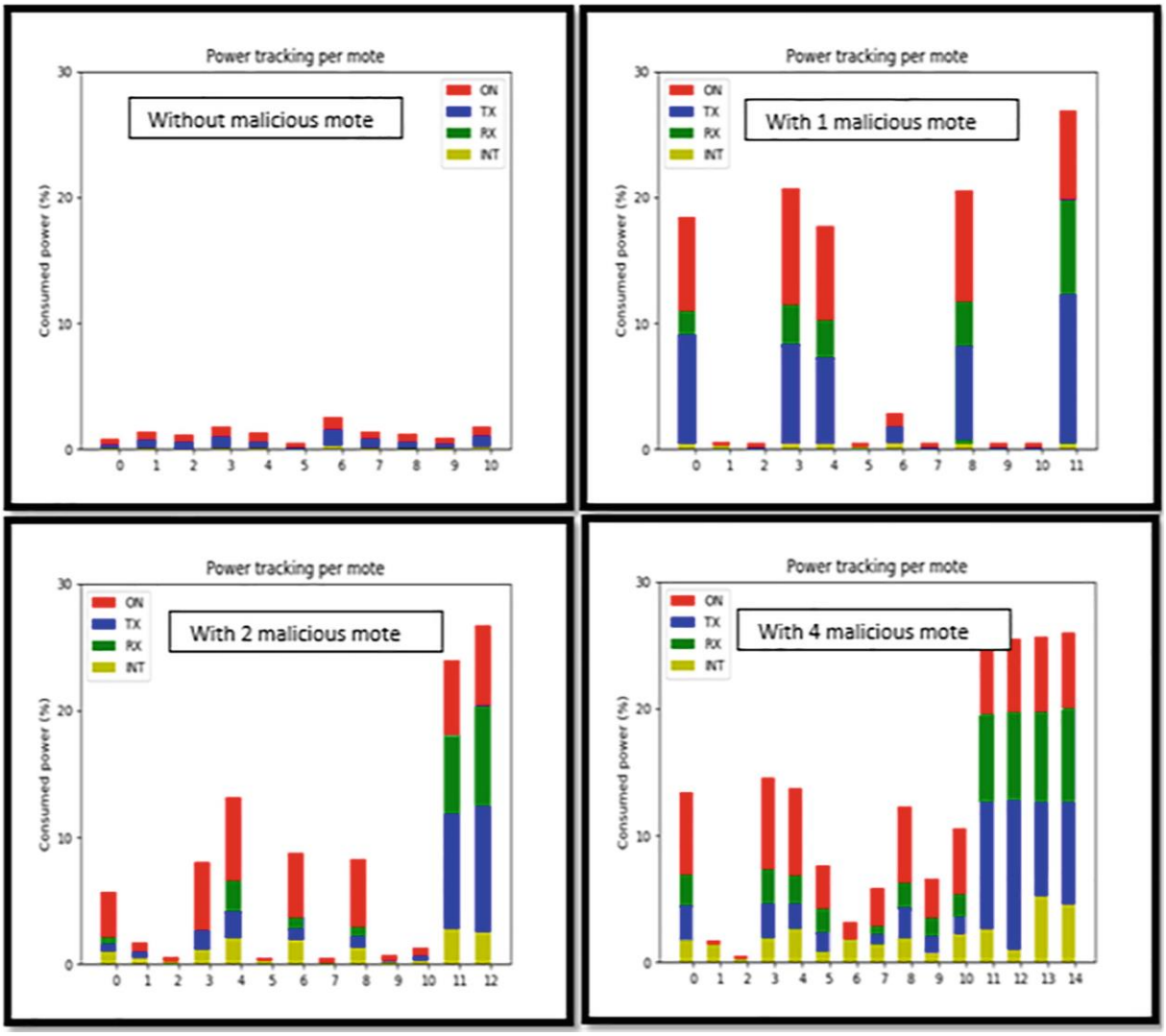


Figure 37 :Suivi de puissance pour chaque mote

Les figures 38 suivantes illustrent les informations obtenues lors de la simulation de l'attaque avec et sans le nœud malveillant, respectivement à droite et à gauche. Une comparaison de la consommation d'énergie été nécessaire pour déduire l'impact de l'attaque en présence et en absence du nœud malveillant.

Nous pouvons remarquer l'impact de l'attaque sur le réseau au fil du temps et en particulier sur les nœuds 3,7 et 10 :

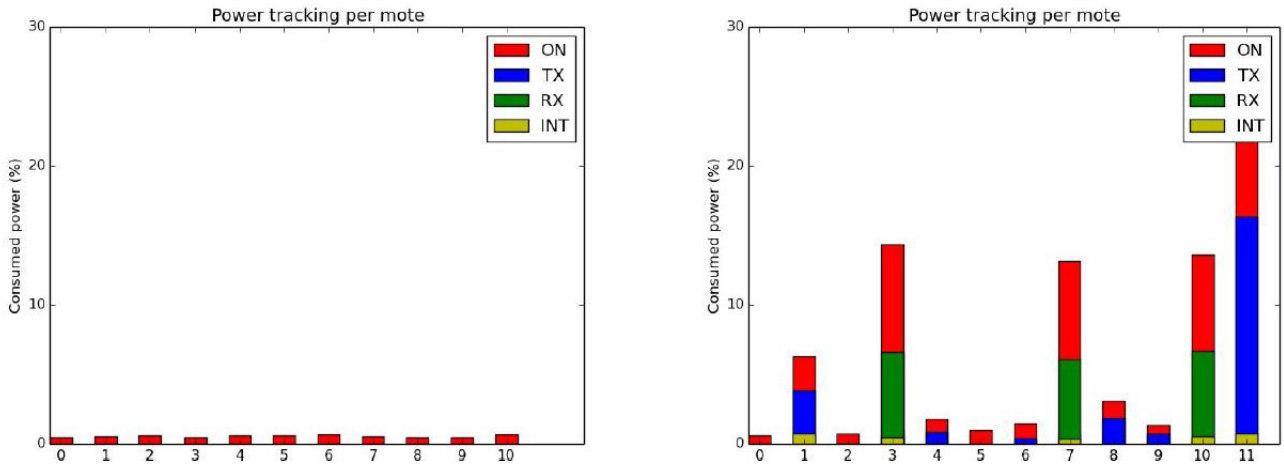


Figure 38:Suivi de puissance sans et avec mote malveillant

On utilisant aussi la Position géographique et le rang (Position sur topologie logique) afin de déterminer la distance entre le nœud et la racine comme un outil de détection des trafics anormal dans notre réseau, elle nous permet de mieux comprendre le comportement des nœuds malveillants (BH) et (DR)

Nous capterons la Position géographique et rang afin d'utiliser aussi comme un mécanisme de sécurité basé sur la force du signal et les informations géographiques pour détecter les nœuds malveillants qui lancent l'attaque. L'idée est de comparer la force du signal d'une réception avec sa valeur attendue (calculée à l'aide des informations géographiques), et de la spécification d'émetteur-récepteur prédéfinie.

Enfin, toutes les données résultant du fichier csv (déjà traite) sont concaténées avec les paramètres d'évaluation de l'énergie et la position géographique avec rang.

La valeur de l'énergie varie selon le temps. Elle a été enregistrée par intermittence. Les données ont été collectées à l'aide d'un capteur pour finalement produire un ensemble de données qui contient toutes les informations nécessaires pour l'apprentissage automatique afin de rendre notre model final efficace contre ces types d'attaque de routage IOT .

7-4 - Normalisation des fonctionnalités :

Les données résultant de différents scénarios d'attaque par routage IoT ont une moyenne et une variance différentes en raison de leur topologie de réseau, ce qui réduit la performance de l'algorithme d'apprentissage automatique. Par conséquent, un processus de normalisation des fonctionnalités est effectué. Nous avons appliqué une transformée de quantification et une mise à l'échelle min-max aux ensembles de données, respectivement. La transformation quantile ajuste la distribution des valeurs des fonctionnalités à la distribution normale. Elle vise à réduire

l'effet négatif des valeurs marginales. Ensuite, nous mettons à l'échelle toutes les valeurs des ensembles de données dans la plage [0-1] par une mise à l'échelle min-max

Les ensembles de données sont normalisés par un processus de normalisation des fonctionnalités afin de rendre le processus de formation plus rapide.

Enfin, toutes les données résultant des différentes topologies de réseau sont concaténées pour produire un ensemble de données pour un type d'attaque de routage IOT.

Nous obtenons en suite trois ensembles de données d'attaque, La collecte de ces ensembles de données construisent notre ensemble des données final par suite nous appliquons des algorithmes d'apprentissages automatiques afin d'arriver à obtenir notre IDS. .

Enfin, toutes les données résultant des différentes topologies de réseau sont concaténées pour produire un ensemble de données pour un type d'attaque de routage IOT. Nous enlevons en suite trois ensembles de données d'attaque, La collecte de ces ensembles de données (qui sont intégrées dans un algorithme d'apprentissage automatique) , est le résultat final (ensemble des données).

Tous ce qui été décrit précédemment pour dégager l'ambigüité de notre contribution est illustré dans le diagramme cité ci-dessous (figure 39).

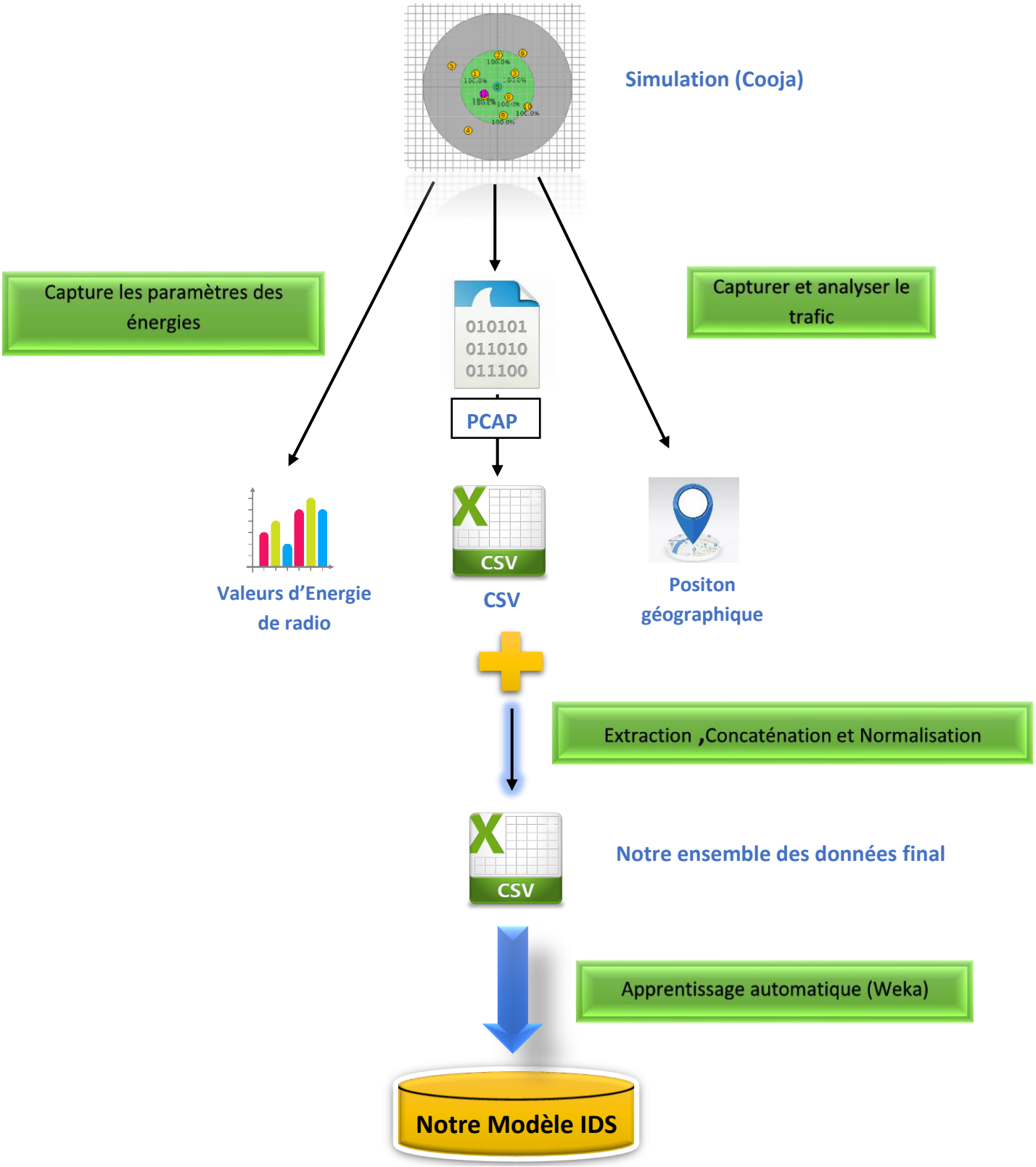


Figure 39 :différentes étapes (phases) du processus de notre modèle

7-5 - Sélections d'attributs :

Nous avons choisi les fonctionnalités de données pour construire notre model d'apprentissage automatique, le choix peut être une cause très essentielle sur les résultats obtenus. Les caractéristiques irréductibles ou partiellement pertinentes peuvent avoir un impact négatif sur les performances du modèle.

N°	Nom de l'attribut	Description
1	T	Temps
2	Src	Source
3	Dst	Destination
4	Protocol	Le protocole de plus haut niveau décodé
5	Dure_tr	Dure de transmission pendant une fenêtre de time
6	Moy_tr	Moyen de transmission
7	Length_tr	La taille du Paquet transmis
8	DIS_tr	Nombre de DIS transmis
9	DIO_tr	Nombre de DIO transmis
10	DAO_tr	Nombre de DAO transmis
11	Dure_rec	Durée de réception pendant une fenêtre de time (1 Second)
12	Moy_rec	Moyen de réception
13	Length_rec	La taille du Paquet reçu
14	DIS_rec	Nombre de DIS reçu
15	DIO_rec	Nombre de DIO reçu
16	DAO_rec	Nombre de DAO reçu
17	ON	Energie d'activité radio
18	TX	Radio d'énergie d'émission
19	RX	Radio d'énergie de réception
20	INT	Radio interférée
21	Pos_x	Position géographique sur l'axe X
22	Pos_y	Position géographique sur l'axe y
23	Rang	Position sur topologie DODAG
24	Class	Classer l'attaque par leur type

Tableau 9: Description de différents attributs

IV - La description du modèle

La plupart des travaux de la détection d'intrusion utilisent les classificateurs du même niveau de façon isolée. Dans cette proposition, nous proposons une approche qui est Représentée dans un modèle de détection d'intrusion hybride et hiérarchique basé sur les travaux de Mr ALEM [62]

Les résultats fournis dans sa recherche étaient excellents, notamment avec la détection des attaques rares où il a utilisé l'ensemble de données bien connu KDD'99, qui représente les données les plus utilisées pour l'IDS.

Du nouveau dans notre proposition c'est l'application de l'idée de Mr ALEM [62] sur l'ensemble de données des attaques IdO que nous avons construit.

Nous avons généré un ensemble de données par des simulations équivalentes à la vie réelle en utilisant le simulateur Contiki/Cooja, , on a aussi porté des modifications au modèle, pour le rendre compatible avec l'environnement IdO.

Ces modifications visent à réduire, de façon efficace, l'échelle de l'ensemble des données afin d'économiser le temps d**'apprentissage** et de la **prévision.**

Notre proposition comprend trois niveaux, dont la particularité réside dans l'intégration des décisions de différents classificateurs du premier niveau par le classificateur du deuxième niveau qui contient le classificateur bayésien naïf qui va confirmer la décision finale du filtre, puis faire le filtrage, afin qu'il ne passe au troisième niveau, que les attaques dont la probabilité de prédiction de Normal est inférieure à 0.6 .

Enfin, le troisième niveau intègre la décision du deuxième niveau haut en fonction de sa prédiction finale, alors que les deux premiers niveaux effectuent une classification de type binaire afin qu'ils arrivent à réduire la complexité des problèmes multi-classes et les rendent plus simples de type classification binaire.

ce troisième niveau est du type multi-classification, et ceci afin de déterminer le type de **l'attaque**, qui est le résultat final, et le niveau deux (02) détermine **le normal** .

Le choix du filtrage normal revient au fait que les instances normales sont beaucoup plus nombreuses que celles anormales, ceci est dû aux caractéristiques du réseau IdO et à la nature de notre ensemble de données.

- **La première raison :** c'est que l'objet IdO à la même importance que le serveur, car si l'attaque touche **l'objet,** le service en générale est **affecté,** et peut tomber si un grand nombre d'objets sont infectées contrairement au réseau : **client-serveur**.

  Voici où réside la difficulté de la sécurité dans le réseau IdO, parce que les **objets** et le **serveur** se **complètent,** et que **l'influence** de **l'un** d'entre **eux** conduit à **la chute du réseau**, suite à l'indisponibilité d'un service lui appartenant, et le rôle de notre proposition est de déterminer tout intrus avant qu'il ne se propage et protège tous les objets et le serveur.

- **La deuxième raison :** est que nous avons adopté notre ensemble des données sur des paquets qui sont groupés par de fenêtres de temps - dans notre cas ΔT=1 seconde

  Quelle que soit l'activité de l'attaque, ceci n'augmentera pas la taille de données du nombre d'instances, mais le changement sera dans le type de données comme indiqué dans le tableau 10 .

Type nœud	ΔT (1 seconde)	Nbre de paquets	Type attaque
Normal	1	136	Normale
malveillante	1	1000	DOS

Tableau 10 :instance avec et sans malveillante

1 - la structure de notre modèle

Dans cette partie, nous présentons le modèle proposé appelé (Minerva-IDS)[1] au cours de notre travail, les différentes étapes ainsi que les résultats qui en découlent. Comme illustrée sur la figure 40, Minerva-IDS se base sur trois niveaux :

- Le premier niveau contient deux meilleurs classificateurs. nous choisissons les meilleurs classificateurs binaires basés sur trois critères principaux :

[1] -L'appellation de notre modèle d'IDS par Minerva-IDS vient de la célèbre chouette mythique de Minerva, cette dernière était associé à la déesse de la raison et de toutes les compétences chez les anciens Romains qui considéraient cette oiseau comme un symbole de sagesse, ou elle était caractérisé par la qualité forte de prendre des décisions judicieuses durant les guerre, pour cela plusieurs artistes la représentaient portant un bouclier magique et un casque, Notre choix de cet oiseau comme symbole de l'IDS est pour plusieurs raisons, dont celles qui distinguent cet oiseau :- Pour son acuité visuelle et son fort sens de l'ouïe,- Pour sa nature de rester éveillée nuit et jour.- Pour sa capacité de tourner complètement sa tête et à 270°, ce qui lui permet d'avoir une meilleur vue.- Pour son efficacité dans la lutte contre tous intrus (rats et souris….).

- Maximiser la précision EXACTITUDE .
- Maximiser le taux de détection DR .
- Minimiser le taux de fausses alertes FAR .

- Le deuxième niveau contient un seul classificateur naïve-bayes , il analyse les prédictions sélectionnées des différents meilleurs classificateurs du premier niveau et prend la décision de filtrage et ne passe au troisième niveau, que les attaques dont la probabilité de prédiction de Normal inférieur à 0,6 .

  Après avoir choisi les meilleurs classificateurs, nous ajoutons les sorties des classificateurs comme autres attributs dans l'ensemble de données de l'apprentissage initiale pour chaque connexion de l'ensemble de données initial, nous donnons la décision Normal ou Anormal qui représentent les sorties des classificateurs sélectionnés.

  Ensuite, l'enregistrement A est représenté par tous les attributs initiaux et deux décisions de C1 et C2 en fonction des attributs initiaux Comme illustré dans la figure 42.

- Le troisième niveau contient le meilleure classificateur multi-classe que nous choisissons basé sur trois critères principaux meilleur taux d'exactitude, taux de détection et le minimum taux de fausses alertes , Où nous fusionnons le résultat du deuxième niveau dans un ensemble de données initial, afin d'augmenter et d'améliorer les résultats, comme indiqué dans le tableau suivant :

Mesures → / **Classificateurs** ↓	DR, %	Exactitude, %	FAR, %	Résultat du deuxième niveau
J48	**0.976**	**0.976**	0.021	Sans
J48	0.996	**0.996**	0.001	Avec

Tableau 11:l'impact de résultat du deuxième niveau

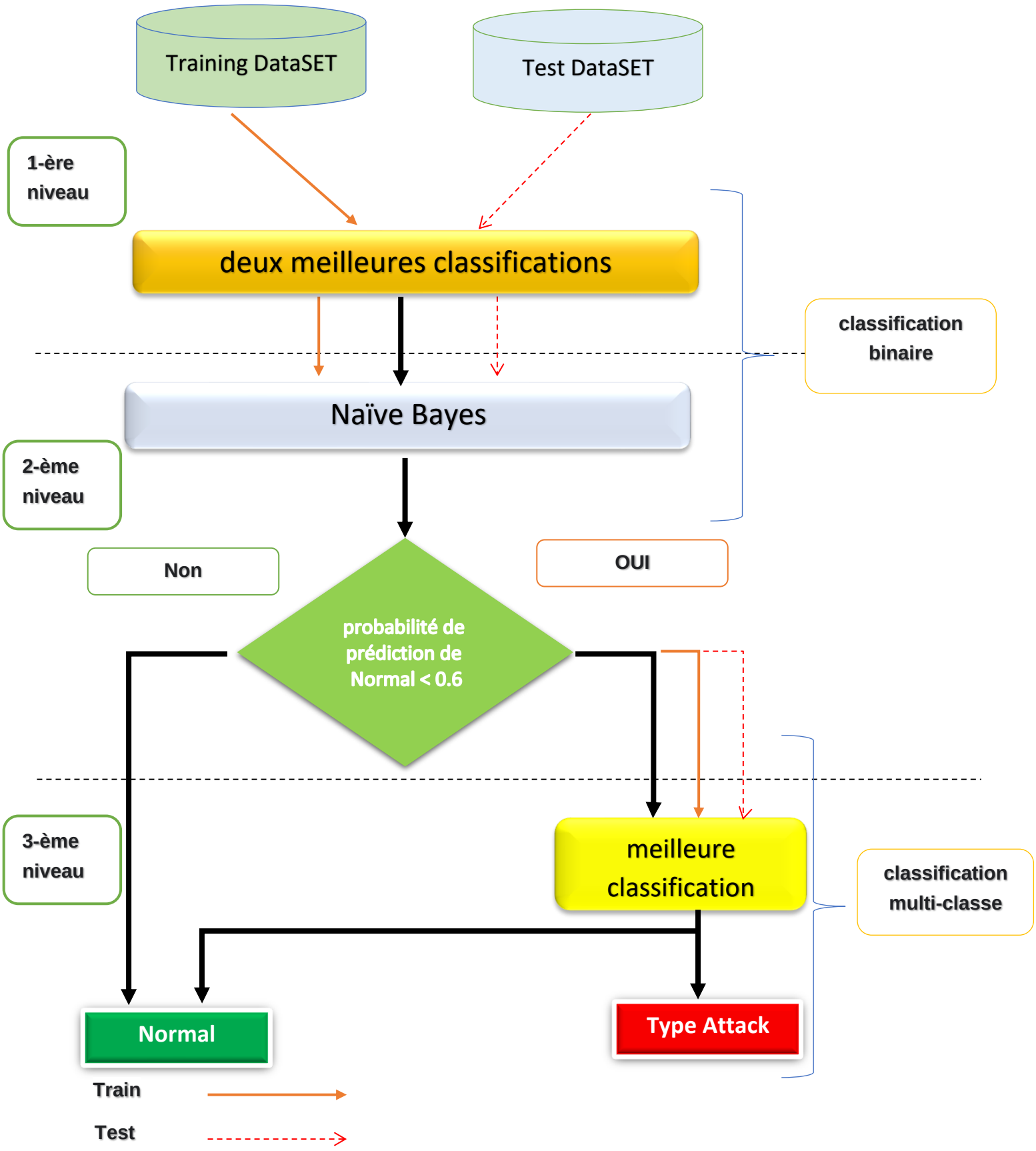


Figure 40:structure générale de Minerva-IDS

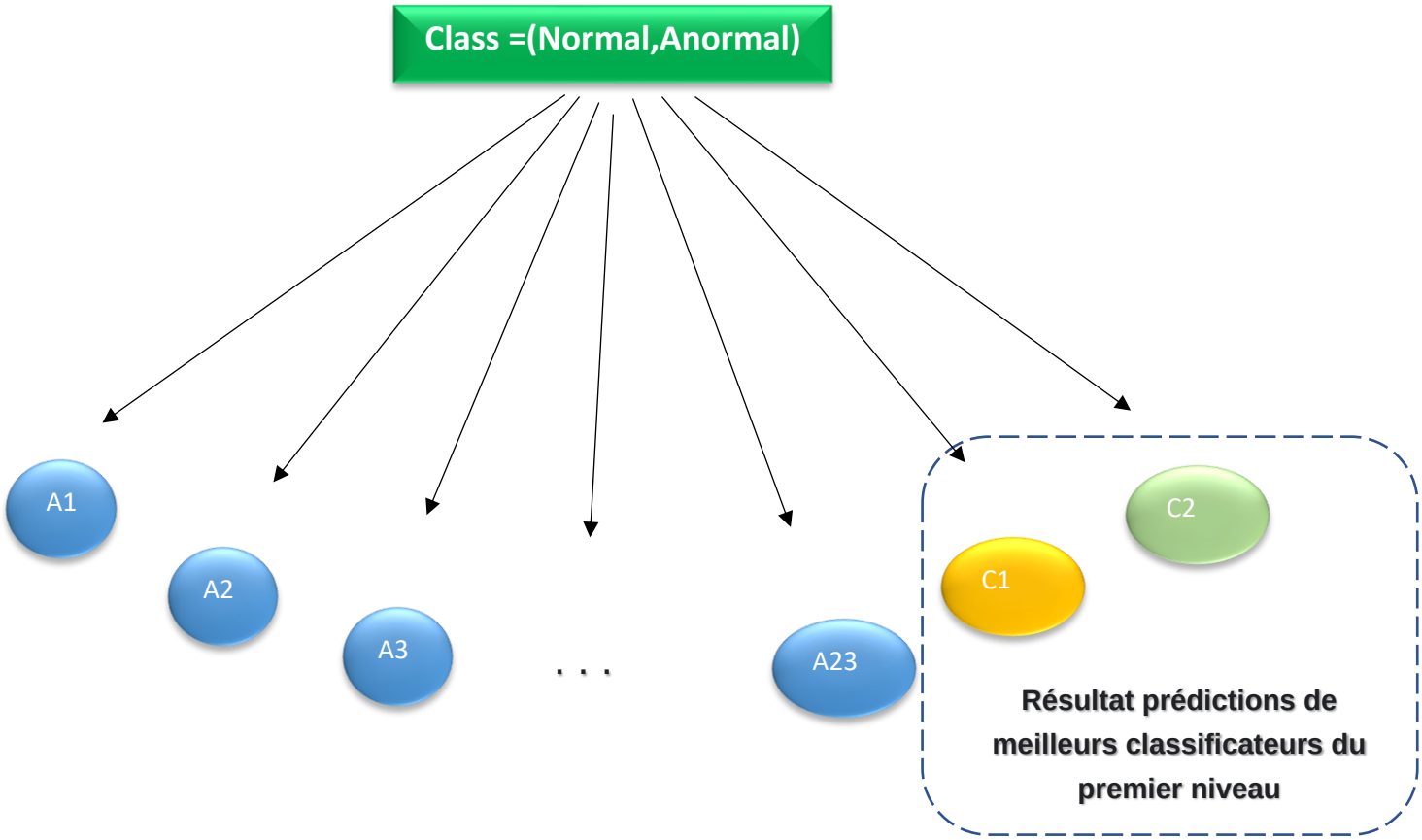


Figure 42:Modèle bayésien (Alem et Al). [62]

2 - Mode de fonctionnement :

La sélection des classificateurs pour la phase **d'apprentissage** et du **test** constituent les étapes qui nous ont permis de construire: Minerva_IDS.

2.1. Sélection des différents classificateurs :

En effet en premier temps, et afin de récolter les meilleurs classificateurs pour Minerva_IDS, on a procédé à une sélection de classificateurs les plus utilisés dans les travaux d'IDS-IdO .

Le **tableau 12** présente une étude comparative entre 04 techniques d'exploration (classification binaire) de données les plus utilisés dans les travaux d'IDS-IdO :

J48, Naïve Bayes **NB**, Random Forest **RF**, LibSVM.

Mesures → / Classificateurs ↓	DR, %	Exactitude, %	FAR, %	Décision
Naivebayesien	0,875	0,856	0,074	
J48	0,999	0,999	0,000	
RandomForest	1,000	1,000	0,000	
LibSVM	0,860	0,860	0,049	

Tableau 12:La comparaison entre les classificateurs(classification binaire)

Ensuite, une élimination de quelques classificateurs parmi ces derniers à comparer chaque classificateur deux a deux sur leur meilleur taux d'exactitude, taux de détection et le minimum taux de fausses alertes, jusqu'à l'obtention de deux meilleurs classificateurs.

C'est ainsi qu'on retiendra les deux classificateurs pour former le premier niveau de Minerva_IDS à savoir : **RandomForest** et **J48**.

Pour le troisième niveau, nous faisons la même opération précédemment faite pour le premier niveau, et nous choisissons le meilleur multi-classificateur .

On comparant les différents classificateurs pour les trois mesures comme illustré dans le tableau 13 :

Mesures → / Classificateurs ↓	DR, %	Exactitude, %	FAR, %	Décision
Naivebayesien	0.958	0.958	0.227	
J48	0,984	0,989	0,009	
RandomForest	0,999	0,999	0,001	
LibSVM	0,967	0,981	0,026	

Tableau 13:La comparaison entre les classificateurs (Multiclasses)

Enfin , nous obtiendrons le meilleur le meilleure classificateur pour former la troisième niveau de Minerva_IDS est : **RandomForest**.

3 - Les ensembles de données d'apprentissage et de test :

Nous avons créé notre ensemble de données contenant 48 024 enregistrements pour l'apprentissage et le test. Cet ensemble contient 80% des données utilisées pour l'apprentissage du modèle , le reste n'est utilisé que pour vérifier et évaluer les performances du modèle.

Le tableau 14 ci-dessous résume la distribution des attaques ainsi le comportement normal de notre ensemble de données d'apprentissage et de test.

	Apprentissage	Test	Total	Taux %
Rank	7493	1874	9367	20%
Version	2556	640	3196	7%
Hello	4036	1010	5046	11%
Black	1194	299	1493	3%
Normal	23125	5797	28922	60%
Total	38404	9620	48024	100%

Tableau 14 : la répartition des attaques et du comportement normal de notre ensemble de données d'apprentissage et de test

4 - Expérimentation :

Nous avons mené une série d'expériences sur notre ensemble de données, qui est l'ensemble de données de détection d'attaque de routage de l'IdO.

Les résultats sont obtenus sur un PC HP Intel(R) Cœur i7-4500U, 2.40 GHz, 8 Go de RAM , et avec le logiciel d'apprentissage automatique WEKA 3.8.4 (Weka est utilisé pour la mise en œuvre des différents classificateurs) .

Ensuite, nous procédons à l'expérimentation du premier niveau, et ceci en utilisant Random Forest, et J48.

Pour le deuxième niveau, nous avons utilisé le classifieur Naïves Bayes, pour la simple raison de sa caractéristique probabiliste, et de sa capacité à intégrer les prédictions avec les classificateurs du premier niveau, et puis nous appliquons le filtre.

Enfin, concernant le troisième niveau, nous avons choisi Random Forest pour émettre un résultat final qui définit le type d'attaque avec les prédictions extraites du classifieur Naive Bayes.

Le tableau 15 qui suit montre les étapes de l'expérience **Minerva-IDS**, avec une description de chaque phase :

	Mesures →	**DR, %**	**Exactitude, %**	**FAR, %**
Etapes	Classificateurs ↓			
Niveau 01	**R**andomForest	1,000	1,000	0,000
	J48	0,999	0,999	0,000
Niveau 02	Naivebayesien	0,999	0,999	0,000
Filtrage	**Taux de probabilité de prédiction de Normal > 0.6 est 60.2 % (qu'il ne passe au troisième niveau)**			
Niveau 03	**R**andomForest	0,999	1,000	0,000
Résultat Final de Minerva IDS (résulta de Niveau 02 + Niveau 03)		0,999	1,000	0,001

Tableau 15:les étapes de l'expérience Minerva-IDS

5 - Etude comparative :

Tout d'abord, nous avons comparé nos résultats avec les résultats des meilleurs classificateurs résumés dans le tableau (16) suivant :

Classifieur	**DR, %**	**Exactitude, %**	**FAR, %**
Naivebayesien	0.958	0.958	0.227
J48	0,984	0,989	0,009
RandomForest	0,999	0,999	0,001
LibSVM	0,967	0,981	0,026
Minerva-Ids	0,999	1,000	0,001

Tableau 16:tableau comparative entre les classificateurs

a = rank	b = version	c = hello	d = black	e = normal	
1874	0	0	0	0	a = rank
0	635	0	4	1	b = version
0	0	1009	0	1	c = hello
0	0	0	299	0	d = black
0	2	0	3	5792	e = normal

Tableau 17 :Matrice de confusion de Random-Forest

a = rank	b = version	c = hello	d = black	e = normal	
1874	0	0	0	0	a = rank
0	640	0	0	0	b = version
0	0	1010	0	0	c = hello
0	0	0	299	0	d = black
1	3	0	1	5792	e = normal

Tableau 18 :Matrice de confusion de Minerva-IDS

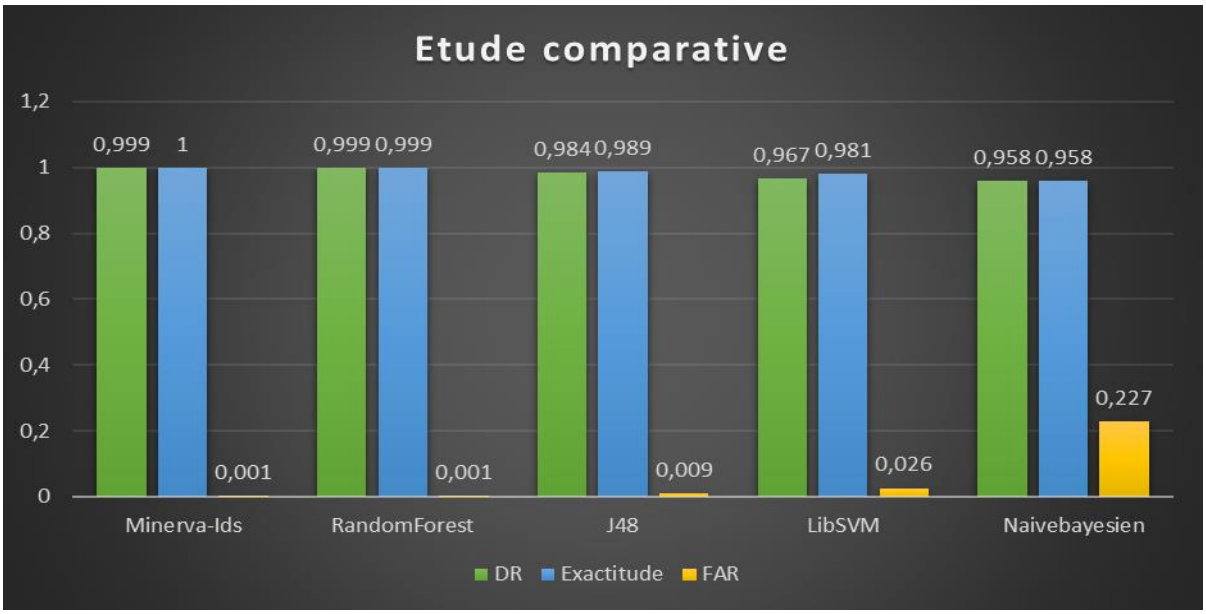


Figure 43:Etude comparative entre les classificateurs

6 - Discussion :

À partir de l'analyse de tableau 16 précédent nous concluons que Minerva-IDS est plus fort que les autres .

Pour évaluer les performances de Minerva_IDS, nous avons comparé ses performances avec des travaux connexes [65]**et** [63] **,**Le résultat de cette étude comparative est résumé dans le tableau suivant.

	Minerva-IDS	[65]	[63]
Simulation	Oui/cooja	Oui/cooja	Oui/cooja
Type attaques	4	3	4
Type data	Pcap files, énergie, position	Pcap files	Pcap files
Nbre d'attributs	23	18	21

Tableau 19 :comparaison des data-set entre les IDS proposés

Classifieur	**Precision**	**Recall**	**F1-Score**	**Accuracy**
[65]	0.957	0.957	0.957	/
[63]	0.994	0.993	/	99.330
Minerva-IDS	0.999	1.000	0.999	1.000

Tableau 20:la comparaison des mesures de performances entre les IDS proposés

Comme le montre le tableau 20, Minerva_IDS a montré sa haute performance pour le taux d'exactitude (Accuracy) le plus élevé, et un fort taux de détection(Recall). En outre, Minerva_IDS est plus précis que les autres utilisés dans cette étude comparative avec un taux d'exactitude égale à 100 % et precision 0.999 ,F1-Score 0.999.

7 - Position de notre modèle :

Lors de notre simulation et d'après notre étude sur la nature de ces attaques, nous avons remarqué que l'attaque va choisir un emplacement stratégique afin d'avoir un grand impact, car plus le Rang des objets est plus grand que le rang de l'attaque et plus éloigné de la racine, plus il est ciblé.

Nous proposons donc une architecture que nous appelons le Cluster-RPL (cluster-star), dont le but sera d'améliorer la sécurité pour réduire l'effet de l'attaque, car nous suggérons que chaque groupe d'objets qui ont le même Rang forme ce qu'on appelle un cluster (groupe) entre eux via le Sous-root, comme montré dans la figure 45 ci-après :

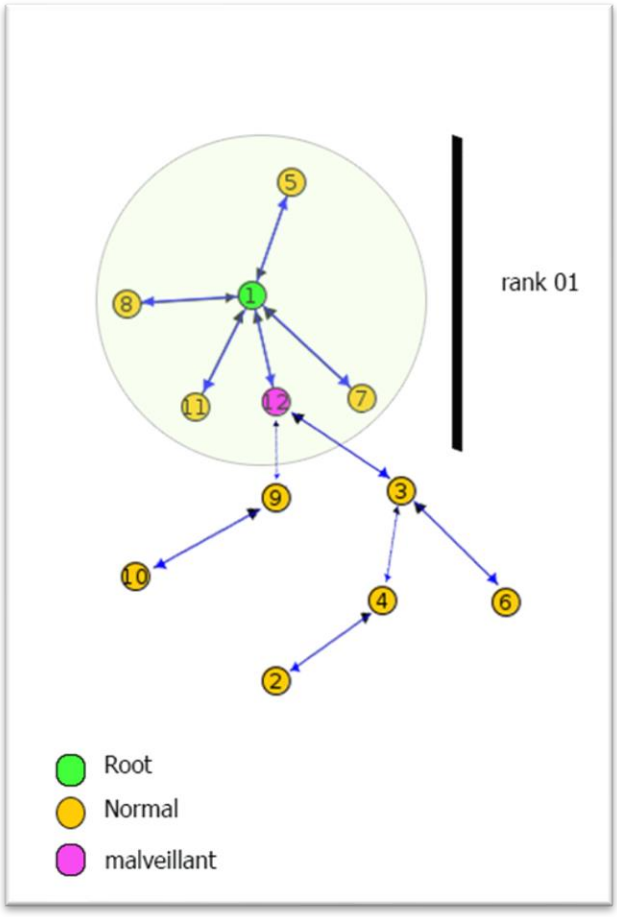


Figure 44 :Topologie RPL (DODAG)

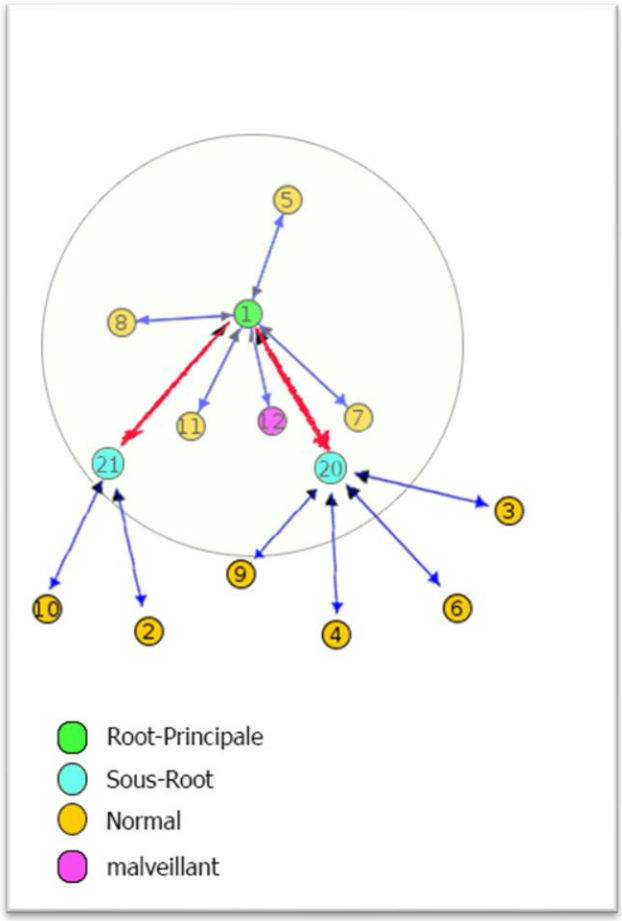


Figure 45:Topologie Cluster-RPL

Avec cette solution, tous les objets auront le même rang , de sorte que le Cluster-RPL des sous-réseaux (des groupes ou des sous-roots) se connectent entre eux et aussi avec le root principal, et c'est ainsi que nous placerons les IDS-Minerva dans chaque sous-racine,

Puisque les IDS représentent la deuxième ligne de défense, il est nécessaire d'avoir un autre mécanisme de sécurité, nous proposons donc d'augmenter cette dernière avec un système d'authentification placé au niveau de chaque racine (sous-root et principal) afin d'identifier les objets du réseau et prévenir tout intrus.

En plus de l'intégration du cryptage IPSec existant dans l'ipv6 , cette solution sera donc peu coûteuse, car la portée moyenne de couverture pour un root dans le cas de l'utilisation de la technologie Wi-Fi , dépend essentiellement de la norme utilisée.

En théorie, la dernière version offre ainsi une portée de 400 mètres en extérieur et de 90 mètres en intérieur, et dans une petite organisation nous n'aurons besoin que de 2 ou 3 sous-roots .

Aussi, l'un des avantages de cette solution est la réduction de la Charge sur les IDS situés au niveau de le root-principal, en raison de la propriété de l'utilisation du sous groupement.

Enfin, cette solution peut être étendue et appliquée sur une zone géographique plus large, telle qu'une ville, et ceci par le biais du **Cloud**, ce qui va permettre aux réseaux RPL de communiquer, de partager, et de traiter les données volumineuses provenant des réseaux hétérogènes.

La sous-racine se limitera à la protection (analyse du trafic par Minerva-IDS), et au routage, et comme cette solution Cluster-RPLdivisera le réseau en petits groupes (cluster), il n'y aura donc pas de consommation d'énergie significative.

Cette solution conduira à une protection renforcée et aussi à déterminer rapidement l'attaque,

et à intervenir s'il y a intrusion, en l'isolant seule, ceci n’empêchera pas le sous arbre (figure 46) dans un cas normal de RPL (si le Cluste-RPLr n'est pas utilisé), comme l’exemple de l’utilisation des ACL’s.

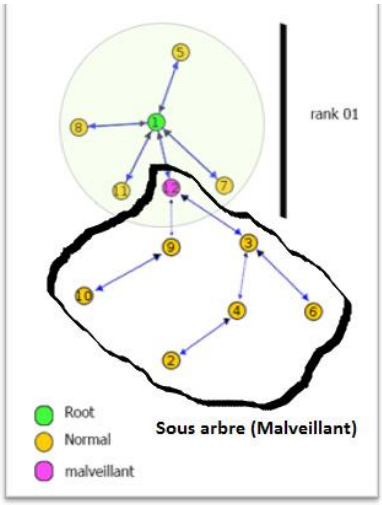


Figure 46 :Sous Arbre (Malveillante)

Conclusion :

Dans ce chapitre, nous avons proposé une solution de détection d'attaque Routage dans un réseau IoT qui vise le RPL comme protocole de routage. Nous avons simulé à l'aide de Contiki-Cooja pas mal de scénarios réseau, pour pouvoir générer et former les jeux de données à utiliser dans la phase de test et d'apprentissage, dans laquelle nous allons utiliser WEKA, pour décider selon la base de données si le comportement est normal ou malveillant.

englobe la démarche suivie et synthétise l'ensemble des résultats obtenus après l'application de notre modèle proposé. Nous employons dans cette proposition, un modèle basé sur la combinaison de 04 classificateurs avec de différentes techniques de classification (binaire ,multiclasses) el filtrage, en collaboration, afin d'améliorer la classification des données.

Comme cité plus haut nous avons montré la haute performance de notre modèle par rapport aux travaux connexes et certains modèles récents de détection d'intrusion.

le Minerva IDS donne un taux d'exactitude le plus élevé avec le taux de fausse alarme le plus bas et un bon taux de réduction dans l'échelle de l'ensemble de l'apprentissage, et réduit considérablement le temps de prédiction

Donc ce chapitre a présenté une étape importante vers le développement d'une architecture globale de détection d'intrusion basée sur l'approche comportementale contre les attaques de routages dans réseau IOT.

Chapitre5 : Minerva-IDS, réalisation et implémentation

Chapitre 05 : Minerva-IDS, réalisation et implémentation

Introduction :

Après avoir présenté l'architecture de notre modèle dans le chapitre précédent et l'illustration que ce système fait de l'amélioration par rapport à d'autres travaux.

Dans ce chapitre nous allons d'abord expliquer les outils utilisés dans la réalisation de notre prototype, l'environnement de développement et le langage de programmation utilisé.

Nous détaillons le processus de l'implémentation ainsi que la principale interface qui le compose à travers des fenêtres de capture.

I - Environnement de programmation :

1 - Présentation de l'environnement JAVA :

Qu'est ce que Java ?

Apparu fin 1995 début 1996 et développé par Sun Microsystems Java s'est très rapidement taillé une place importante en particulier dans le domaine de l'internet et des applications client-serveur. Les objectifs de Java sont d'être multiplateformes et d'assurer la sécurité aussi bien pendant le développement que pendant l'utilisation d'un programme Java. Il est en passe de détrôner le langage C++ dont il hérite partiellement la syntaxe mais non ses défauts. Comme C++ et Delphi, Java est algorithmique et orienté objet à ce titre il peut effectuer comme ses compagnons, tout les taches d'un tel langage (bureautiques, graphiques, multimédias, base de données, environnement de développement, etc….) .son point de fort qui le démarque des autres est sa portabilité due(en théorie) à ses bibliothèques de classes indépendantes de la plate-forme ,ce qui est le point essentiel de la programmation sur internet ou plusieurs machines dissemblables sont interconnectées. [73]

2 - Présentation de NetBeans IDE :

NetBeans est un environnement de développement intégré (EDI) a été créé à l'initiative de Sun Microsystems. Il présente toutes les caractéristiques indispensables à un environnement de qualité, que ce soit pour développer en Java, Ruby, C/C++ ou même PHP.

NetBeans est sous licence OpenSource, il permet de développer et déployer rapidement et gratuitement des applications graphiques Swing, des Applets, des JSP/Servlets, des architectures J2EE, dans un environnement fortement personnalisable.

L'IDE NetBeans repose sur un noyau robuste, la plateforme NetBeans, que vous pouvez également utiliser pour développer vos propres applications Java, et un système de plugins performant, qui permet d'avoir un IDE modulable.

Enfin cet IDE possède un débogueur de grande qualité ainsi qu'une interface graphique améliorée.[74]

3 - Présentation de Weka :

Weka (Waikato Environment for Knowledge Analysis) est une suite populaire de logiciels d'apprentissage automatique. Écrite en Java, développée à l'université de Waikato, Nouvelle-Zélande. Weka est un Logiciel libre disponible sous la Licence publique générale GNU (GPL). L'espace de travail Weka contient une collection d'outils de visualisation et d'algorithmes pour l'analyse des données et la modélisation prédictive, allié à une interface graphique pour un accès facile de ses fonctionnalités donc il se compose principalement[75]

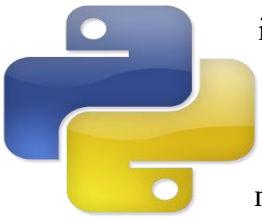
- De classes Java permettant de charger et de manipuler les données.
- De classes pour les principaux algorithmes de classification supervisée ou non supervisée.
- D'outils de sélection d'attributs, de statistiques sur ces attributs.
- De classes permettant de visualiser les résultats.

On peut l'utiliser à trois niveaux :

- Via l'interface graphique, pour charger un fichier de données, lui appliquer un algorithme, vérifier son efficacité.

- Invoquer un algorithme sur la ligne de commande.
- Utiliser les classes définies dans ses propres programmes pour créer d'autres méthodes, implémenter d'autres algorithmes, comparer ou combiner plusieurs méthodes.

4 - Définition du langage Python en informatique :



Python est le langage de programmation open source le plus employé par les informaticiens. Ce langage s'est propulsé en tête de la gestion d'infrastructure, d'analyse de données ou dans le domaine du développement de logiciels. En effet, parmi ses qualités, Python permet notamment aux développeurs de se concentrer sur ce qu'ils font plutôt que sur la manière dont ils le font. Il a libéré les développeurs des contraintes de formes qui occupaient leur temps avec les langages plus anciens. Ainsi, développer du code avec Python est plus rapide qu'avec d'autres langages. [76]

5 - Définition jupyter :

Jupyter se présente comme un outil extrêmement simple à mettre en œuvre qui vous permettra de transformer vos Jupyter Notebooks en applications web ou en Dashboard quasiment automatiquement.[77]

6 - Bibliothèques Supplémentaires :

Afin d'atteindre les objectifs de ce projet, nous avons utilisé d'autres bibliothèques externes pour effectuer certaines tâches spécifiques

6-1 - Pandas :

Est une librairie Python qui a pour objectif de vous faciliter la vie en matière de manipulation de données. Les structures de données gérées par Pandas peuvent contenir tout type d'éléments à savoir (dans le jargon Pandas) des Séries et Data Frame et des Panel. Dans le cadre de nos expérimentations on utilisera plutôt les Data frame car ils offrent une vue bidimensionnelle des données (comme un tableau Excel), et c'est exactement ce que l'on va chercher à utiliser pour nos modèles. [78]

II - Les étapes de la réalisation du projet

Dans notre travail, nous allons créer un prototype qui résume les étapes faites pour atteindre à notre modèle, montrer la différence entre la prédiction de notre approche et celle des classificateurs utilisés dans les différents niveaux de cette dernière.

On lance l'application, la première interface est une fenêtre de login.

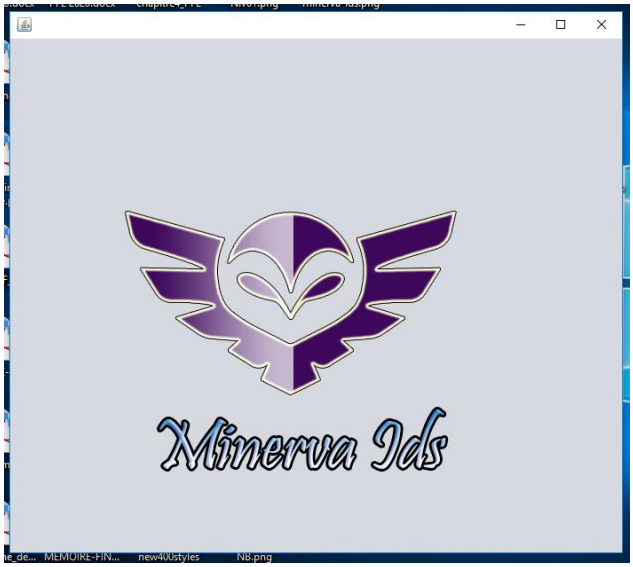


Figure 47:interface de login

Ensuite, le scénario de l'exécution se lancera comme suit :

Dans cette fenêtre, nous pouvons voir la présence de 05 onglets :

Le premier onglet est chargé d'obtenir le corpus de test, il affiche également des informations sur le dataset (type de classificateur et comptage), et contient aussi un bouton d'exécution de Minerva-IDS comme le montre la figure 48 suivante :

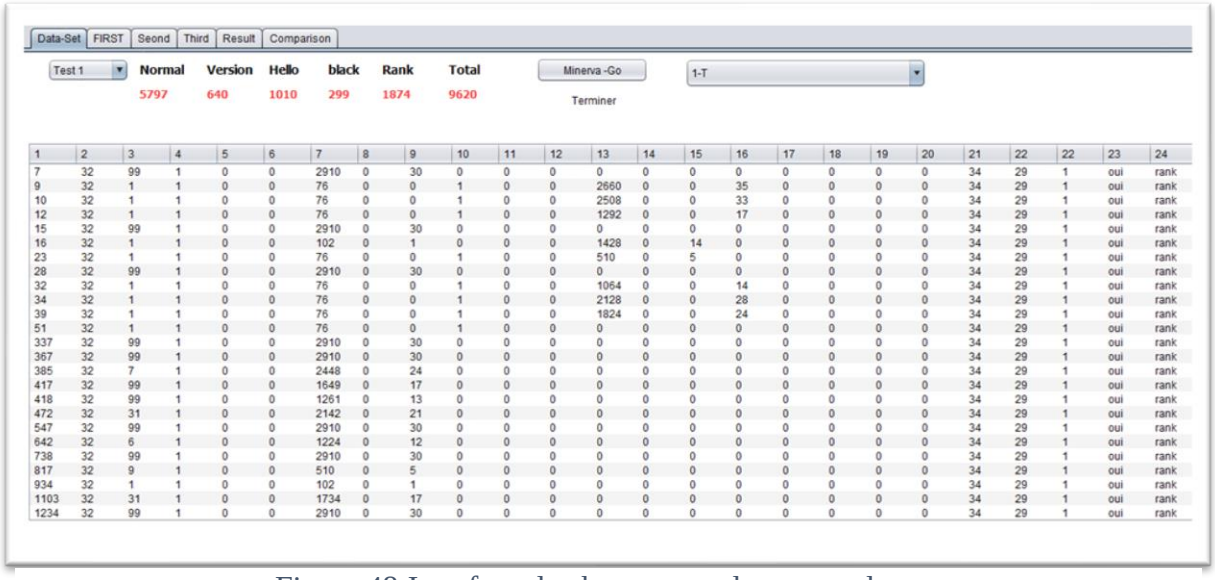


Figure 48:Interface de chargement du corpus de test

Le deuxième onglet contient les classificateurs du premier niveau : RF et J48, et affiche les informations d'évaluation de chaque classificateur à savoir : le DR, l'exactitude, le FAR et la matrice de confusion.

Figure 49:Interface de niveau 01

Le troisième onglet contient un classificateur du niveau 2 (bayes naïfs) où apparaît aussi : le DR, l'exactitude et le FAR avec la présence de la matrice de confusion également.

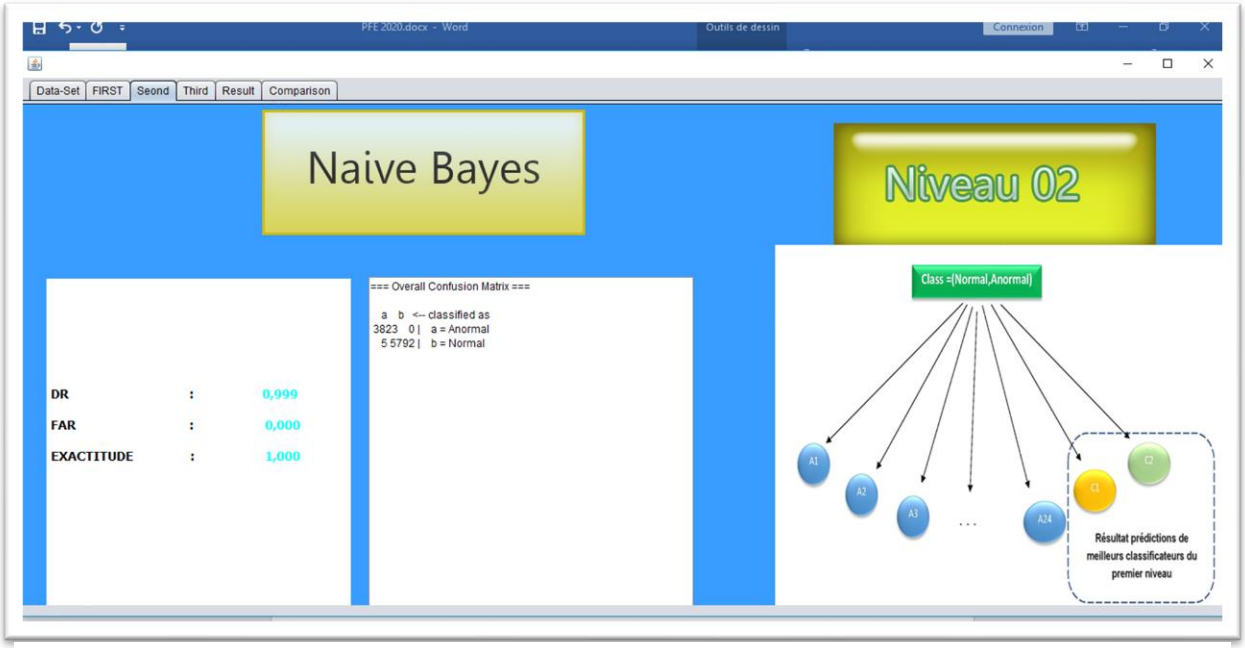


Figure 50:Interface de niveau 02

Le quatrième onglet représente le niveau 03, où se trouve le meilleur classificateur multi-classe RF (Random Forest), avec la présence des mêmes informations que les deux onglets précédents concernant son évaluation.

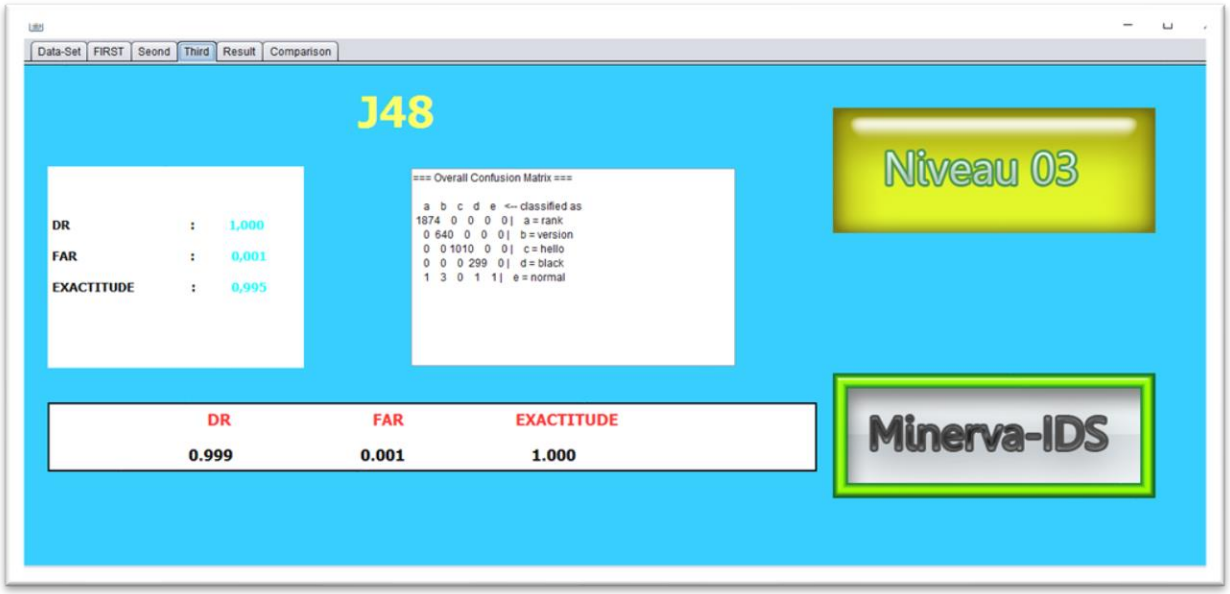


Figure 52 :Interface de niveau 03

Quant au cinquième onglet, il montre les résultats des Minerva-ids en détail, avec la présence d'un bouton pour montrer le pourcentage du filtrage comme indiqué dans la figure 53 qui suit :

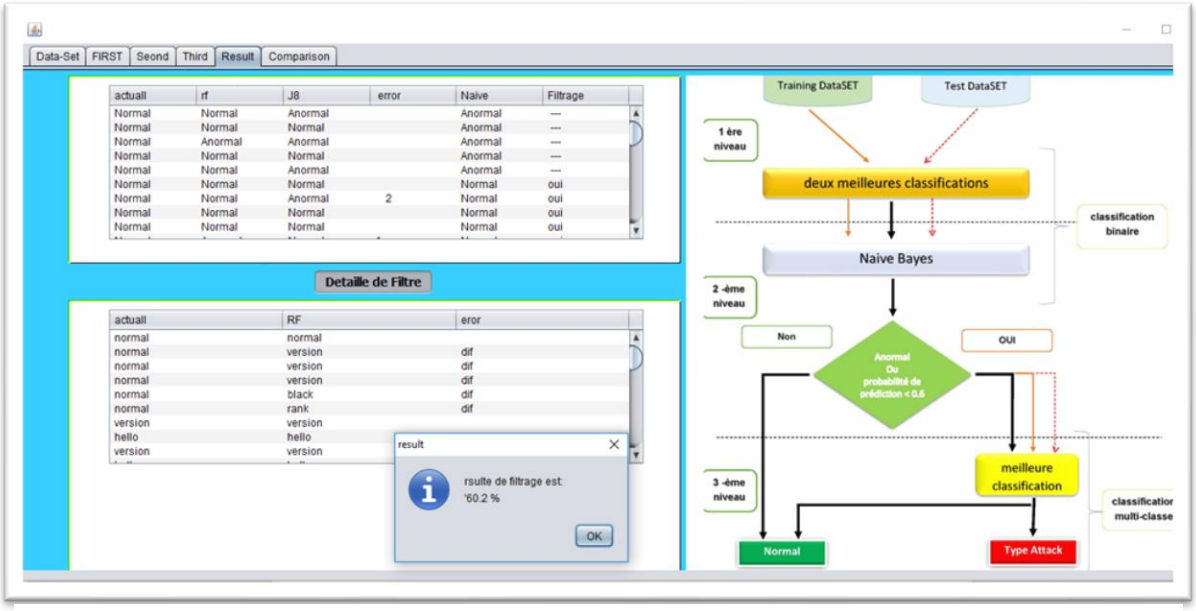


Figure 53:Interface de détail

Enfin, une fenêtre contenant les résultats de la comparaison des Minerva-IDS avec les résultats des travaux antérieurs, en ajoutant des algorithmes de classification.

L'interface suivante présente une comparaison entre les résultats des classificateurs.

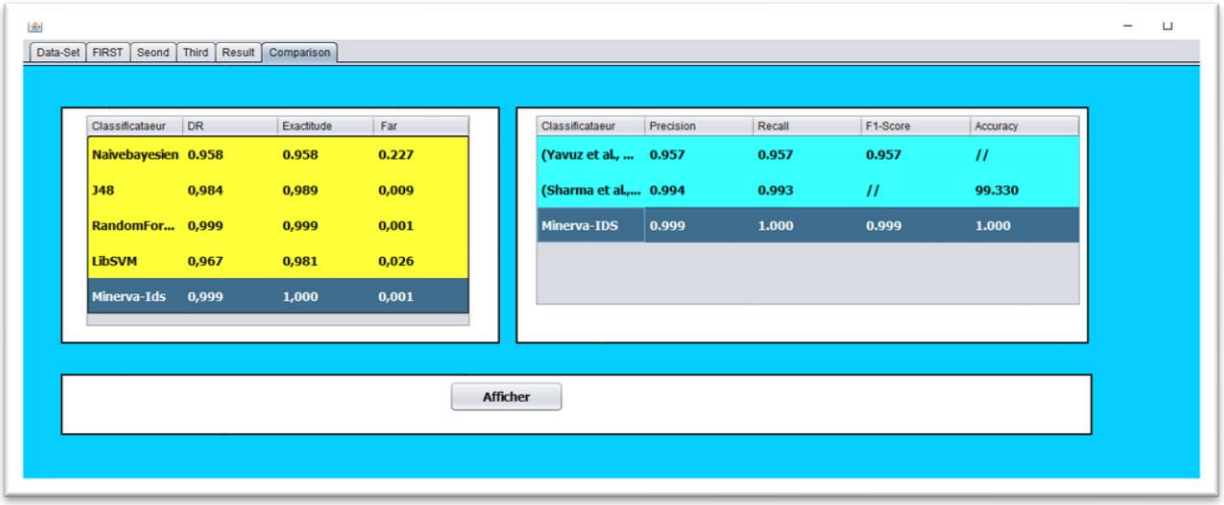


Figure 54:Interface de comparaison

Conclusion

Dans ce chapitre on a présenté quelques outils nécessaires à la réalisation de notre application (l'environnement Netbeans, WEKA,…) et quelques interfaces de notre prototype de l'IDS proposé afin de démontrer la capacité de notre système et en fusionner les 4 classificateurs, et ce pour aboutir à :

- la minimisation du taux de fausses alertes.
- Maximiser le taux de détection.
- réduit considérablement le temps d'apprentissage et de prédiction.

Conclusion Générale

Conclusion Général :

Dans ce travail, nous avons étudié l'impact des attaques de routage (RPL) dans un réseau IdO, afin d'arriver à construire un IDS efficace pour cet environnement. On a aussi analysé la nature des attaques et le comportement normal du réseau IdO pour obtenir un bon IDS.

Notre étude a pour but de collecter les informations capturées durant les communications simulées entres les nœuds d'un réseau IdO et le nœud malveillant dans un réseau RPL en se basant sur l'apprentissage automatique.

Nous avons pu déterminer, à l'aide du simulateur Contiki Cooja plusieurs scénarios du réseau . Ensuite, nous avons construit notre ensemble des données, qui est nécessaire pour la phase d'apprentissage, en utilisant des paramètres importants pour détecter les attaques de différentes taxonomie des attaques de routage (Blackhole et hello-flood et decrease-rank-attack, et version number modification), tels que la fraction de livraison et de réception de paquets des messages de contrôle DODAG (DIO, DIS et DAO) échangés et les valeurs radio-énergie (ON, RX, TX, INT) avec la position géographique (GPS) et la position topologie (range) .

Après avoir obtenu notre ensemble de données, nous avons construits notre modèle hybride et hiérarchique de trois niveaux appelé Minerva-IDS qui combine entre les classificateurs binaires et les multi-classificateurs afin de réduire la taille de l'ensemble de données, et minimiser un maximum de fausses alertes.

Les résultats obtenus : un model robuste et efficaces contre les attaques de routage dans l'ido avec une bonne réduction du temps de prédiction, en comparant nos résultats à des travaux antérieurs dans le même domaine de notre étude ces derniers étaient meilleurs grâce aux informations d'évaluation à savoir : le DR, l'exactitude, le FAR.

Nous espérons approfondir cette étude et proposons un mécanisme de sécurité touchant à l'architecture de routage RPL nommée cluster-RPL afin de renforcer la sécurité des réseaux IDO. .

Enfin, bien que les objectifs de notre recherche soient satisfaisants, nous devons souligner que cet humble travail n'est qu'une simple expérience pour répondre à l'un des problèmes rencontrés dans les attaques de routage dans l'IdO.

A l'avenir, nous essayerons de concrétiser nos modèles Minerva-IDS et RPL-CLUSTER afin de pouvoir détecter de nouvelles attaques et ce dans le but de les rendre plus performants et avoir un ensemble de données référentiel des attaques de routage IdO.

Bibliographie

[1] B. Cousin, “Sécurité des réseaux informatiques.”

[2] G. Pujolle and O. Salvatori, *Cours réseaux et télécoms: avec exercices corrigés*. Eyrolles, 2008.

[3] D. Burgermeister and J. Krier, “Les systèmes de détection d’intrusions.” Article disponible sur http://dbprog. developpez. com, 2006.

[4] B. Ulmann, “Cisco et la sécurité.” Novembre, 2004.

[5] R. Rhouma, “Audit et Sécurité Informatique.” https://sites.google.com/site/rhoouma/teaching-at-esen/cryptographie-et-securite-de-l-information.

[6] Victor MORARU, “La sécurité dans les réseaux haut débit,” pp. 1–33, 2005.

[7] W. Stallings, *Cryptography and network security, 4/E*. Pearson Education India, 2006.

[8] J. Chaouki, S. Michaël, and H. Anis, “TER Détection d’anomalies sur le réseau,” *Rapp. Proj. Univ. Paris Descartes*, 2009.

[9] C. Llorens, L. Levier, D. Valois, and B. Morin, *Tableaux de bord de la sécurité réseau*. Editions Eyrolles, 2011.

[10] L. Bloch, C. Wolfhugel, C. Queinnec, H. Schauer, and N. Makarévitch, *Sécurité informatique: Principes et méthodes à l’usage des DSI, RSSI et administrateurs*. Editions Eyrolles, 2013.

[11] E. Berthomier, “Formation Sécurité des Réseaux,” *Mars*, 2005.

[12] G. Desgeorge, “La sécurité des réseaux.” Cour, 2000.

[13] “IDS.” http://wapiti.enic.fr/commun/ens/peda/options/ST/RIO/pub/exposes/exposesrio2005ttnfa2006/debock-marcant/histo.html (accessed Feb. 06, 2020).

[14] I. DABOUR and I. HADJI, “Etude et mise en place d’un système de détectionprévention d’intrusion (IDSIPS) réseau. Etude de cas SNORT.” .

[15] J.-M. Percher, R. Puttini, L. Mé, O. Camp, B. Jouga, and P. Albers, “Un système de détection d’intrusions distribué pour réseaux ad hoc,” *Tech. Sci. informatiques*, vol. 23, no. 3, pp. 391–420, 2004, doi: 10.3166/tsi.23.391-420.

[16] “CIDF.” http://gost.isi.edu/cidf/ (accessed Feb. 06, 2020).

[17] D. Curry, “Intrusion detection message exchange format data model and extensible mark-up language (xml) document type definition,” *Draft. txt*, 2002.

[18] A. Lukatsky, *Protect your information with intrusion detection*. БХВ-Петербург, 2002.

[19] “Les sondes de sécurité IDS/IPS.” http://igm.univ-mlv.fr/~dr/XPOSE2009/Sonde_de_securite_IDS_IPS/IPS.html (accessed Feb. 06, 2020).

[20] M. Dacier, H. Debar, and A. Wespi, “A Revised Taxonomy for Intrusion-Detection Systems,” *Techmical Rep. Comput. Sci. Math.*, 1999.

[21] M. G. El Rab, “Evaluation des systèmes de détection d’intrusion.” 2008.

[22] C. Bidan, G. Hiet, L. Mé, B. Morin, and J. Zimmermann, “Vers une détection d’intrusions à fiabilité et pertinence prouvables,” *Rev. L Electr. L Electron.*, vol. 9, p. 75, 2006.

[23] F. Majorczyk, “Détection d’intrusions comportementale par diversification de COTS: application au cas des serveurs web.” 2008.

[24] “Les systèmes de détections d’intrusions.” mrproof.blogspot.com/2010/11/les-systemes-de-detections-dintrusions.html (accessed Feb. 06, 2020).

[25] A. Patel, M. Taghavi, K. Bakhtiyari, and J. C. JúNior, “An intrusion detection and prevention system in cloud computing: A systematic review,” *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 25–41, 2013.

[26] B. B. Zarpelão and R. S. Miani, “Detection in I nternet of Things, Journal of Network and Computer Applications,” 2017.

[27] R. Fu, K. Zheng, D. Zhang, and Y. Yang, “An intrusion detection scheme based on anomaly mining in Internet of Things,” 2011.

[28] S. M. Bellovin, “Distributed firewalls.” login, 1999.

[29] “Les systèmes de détection d’intrusions.” https://dbprog.developpez.com/securite/ids/ (accessed Feb. 06, 2020).

[30] J. Yann Berthier and Baptiste, “Détection d’intrusions et analyse forensique,” 2004.

[31] A. Ahmed, “Système de détection d’intrusion adaptatif et distribué,” badji mokhtar annaba, 2014.

[32] P.-J. Benghozi, S. Bureau, and F. Massit-Folléa, “L’Internet des objets/The Internet of Things.” Paris, Editions de la Maison des Sciences de l’Homme, coll. praTICs, 2009.

[33] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of things: A survey on enabling technologies, protocols, and applications,” *IEEE Commun. Surv. tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[34] J. Vasseur, N. Agarwal, J. Hui, Z. Shelby, P. Bertrand, and C. Chauvenet, “RPL: The IP routing protocol designed for low power and lossy networks,” *Internet Protoc. Smart Objects Alliance*, vol. 36, 2011.

[35] M. R. Palattella *et al.*, “Standardized protocol stack for the internet of (important) things,” *IEEE Commun. Surv. tutorials*, vol. 15, no. 3, pp. 1389–1406, 2012.

[36] P. Sethi and S. R. Sarangi, “Internet of things: architectures, protocols, and

applications," *J. Electr. Comput. Eng.*, vol. 2017, 2017.

[37] I. Mashal, O. Alsaryrah, T.-Y. Chung, C.-Z. Yang, W.-H. Kuo, and D. P. Agrawal, "Choices for interaction with things on Internet and underlying issues," *Ad Hoc Networks*, vol. 28, pp. 68–90, 2015.

[38] Y. Sennoun, "IoT & Les protocoles de communication pour les réseaux sans-fil et filaires : Comment choisir ?" https://blog.engineering.publicissapient.fr/2018/08/29/iot-les-protocoles-de-communication-pour-les-reseaux-sans-fil-et-filaires-comment-choisir/ (accessed Jun. 01, 2020).

[39] J. Ko, A. Terzis, S. Dawson-Haggerty, D. E. Culler, J. W. Hui, and P. Levis, "Connecting low-power and lossy networks to the internet," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 96–101, 2011.

[40] I. S. Association, "IEEE Std 802.15. 4-2011, IEEE standard for local and metropolitan area networks—part 15.4: Low-rate wireless personal area networks (LR-WPANs)." Sep, 2011.

[41] E. C. Jones and C. A. Chung, *RFID and Auto-ID in Planning and Logistics: A Practical Guide for Military UID Applications*. CRC Press, 2016.

[42] D. Minoli, *Building the internet of things with IPv6 and MIPv6: The evolving world of M2M communications*. John Wiley & Sons, 2013.

[43] C. Gomez and J. Paradells, "Wireless home automation networks: A survey of architectures and technologies," *IEEE Commun. Mag.*, vol. 48, no. 6, pp. 92–101, 2010.

[44] F. B. Y.ait mouhoub, "Propotion d'un modèle de confaince pour l'internet des Objets," Université A/MIRA de Bejaia, 2015.

[45] T. Limbasiya and N. Doshi, "An analytical study of biometric based remote user authentication schemes using smart cards," *Comput. Electr. Eng.*, vol. 59, pp. 305–321, 2017.

[46] J. Nassar *et al.*, "Fonction objectif pour un RPL adapté aux Smart Grids To cite this version : HAL Id : hal-01513187 Fonction objectif pour un RPL adapté aux Smart Grids," 2017.

[47] T. Winter *et al.*, "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks.," *rfc*, vol. 6550, pp. 1–157, 2012.

[48] A. Mayzaud, R. Badonnel, and I. Chrisment, "A Taxonomy of Attacks in RPL-based Internet of Things," 2016.

[49] K. Weekly and K. Pister, "Evaluating sinkhole defense techniques in RPL networks," in *2012 20th IEEE International Conference on Network Protocols (ICNP)*, 2012, pp. 1–6.

[50] A. Le, J. Loo, Y. Luo, and A. Lasebae, "The impacts of internal threats towards routing protocol for low power and lossy network performance," in *2013 IEEE Symposium on Computers and Communications (ISCC)*, 2013, pp. 789–794.

[51] A. Mayzaud, A. Sehgal, R. Badonnel, I. Chrisment, and J. Schönwälder, “A study of RPL DODAG version attacks,” in *IFIP international conference on autonomous infrastructure, management and security*, 2014, pp. 92–104.

[52] K. Chugh, L. Aboubaker, and J. Loo, “Case study of a black hole attack on LoWPAN-RPL,” in *Proc. of the Sixth International Conference on Emerging Security Information, Systems and Technologies (SECURWARE), Rome, Italy (August 2012)*, 2012, pp. 157–162.

[53] L. Blanc, “La sécurité de l’Internet des Objet.”

[54] R. H. Weber, “Internet of Things–New security and privacy challenges,” *Comput. law Secur. Rev.*, vol. 26, no. 1, pp. 23–30, 2010.

[55] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, “Internet of things: Vision, applications and research challenges,” *Ad hoc networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[56] K. Gaurav, P. Goyal, V. Agrawal, and S. L. Rao, “IoT transaction security,” in *Proceedings of the 5th International Conference on the Internet of Things (IoT), Seoul, Korea*, 2015, pp. 26–28.

[57] P.-J. Benghozi, S. Bureau, and F. Massit-Folea, “L’Internet des objets. Quels enjeux pour les Européens?,” 2008.

[58] R. Greenstadt and J. Beal, “Cognitive security for personal devices,” in *Proceedings of the 1st ACM workshop on Workshop on AISec*, 2008, pp. 27–30.

[59] “Qu’est-ce que l’edge computing?” https://justaskthales.com/fr/quest-ce-que-ledge-computing/.

[60] “Edge computing: traiter la data à la source.” https://www.ictjournal.ch/articles/2020-02-06/edge-computing-traiter-la-data-a-la-source (accessed Feb. 06, 2020).

[61] T. DARIEL, “Introduction à l’IOT (Internet des objets),” 2018.

[62] A. Alem, Y. Dahmani, and B. Mebarek, “Skyline computation for improving naïve Bayesian classifier in intrusion detection system,” *Ing. des Syst. d’Information*, vol. 24, no. 5, pp. 513–518, 2019, doi: 10.18280/isi.240508.

[63] M. Sharma, H. Elmiligi, F. Gebali, and A. Verma, “Simulating Attacks for RPL and Generating Multi-class Dataset for Supervised Machine Learning,” *2019 IEEE 10th Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2019*, pp. 20–26, 2019, doi: 10.1109/IEMCON.2019.8936142.

[64] A. Yahyaoui, F. Yaakoubi, and T. Abdellatif, “Machine Learning Based Rank Attack Detection for Smart Hospital Infrastructure,” in *International Conference on Smart Homes and Health Telematics*, 2020, pp. 28–40.

[65] F. Y. Yavuz, D. Ünal, and E. Gül, “Deep learning for detection of routing attacks in the internet of things,” *Int. J. Comput. Intell. Syst.*, vol. 12, no. 1, pp. 39–58, 2018, doi:

10.2991/ijcis.2018.25905181.

[66] Y. Yao *et al.*, “K-SVM: An effective SVM algorithm based on K-means clustering,” *J. Comput.*, vol. 8, no. 10, pp. 2632–2639, 2013, doi: 10.4304/jcp.8.10.2632-2639.

[67] “Cooja Simulator.” https://anrg.usc.edu/contiki/index.php/Cooja_Simulator (accessed Jul. 10, 2020).

[68] “Apprentissage Supervisé.” https://machinelearnia.com/apprentissage-supervise-4-etapes/?fbclid=IwAR2TI466K-dZSzo8ljjvX5pob1TsojLg_UMFp1X1g6aM3BW7TPCkug2SMeE (accessed Oct. 16, 2020).

[69] “Apprentissage Non-Supervisé.” https://www.lemagit.fr/definition/Apprentissage-non-supervise.

[70] Adrien Haccoun, “Comparaison de méthodes de classifications.” [Online]. Available: https://www.lri.fr/~antoine/Courses/Master-ISI/ISI-10/Projets_2012/Projet_DM.pdf.

[71] D. A, “Classi cation Arbres de décision,” 2015.

[72] O. PARENT and J. EUSTACHE, “Les réseaux bayésiens,” *Univ. Claude Bernard Lyon*, vol. 1, 2006.

[73] “JAVA.” https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203555-java-definition/.

[74] “Netbeans IDE.” https://netbeans.org/features/java/index.html.

[75] “Weka.” https://www.cs.waikato.ac.nz/ml/weka/?fbclid=IwAR2ZaIuZyVXM-v51C3YvCs7NMnuxGLJTeEoHEq67KdxXX3bU9ByLuJqxnhs (accessed Oct. 16, 2020).

[76] “Python.” https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1445304-python-definition-et-utilisation-de-ce-langage-informatique/.

[77] “Jupyter.” https://jupyter.org/.

[78] “Pandas.” https://pandas.pydata.org/.

[79] P. Porras, D. Schnackenberg, S. Staniford-Chen, M. Stillman, and F. Wu, “The common intrusion detection framework architecture (CIDF),” *Univ. Calif.*, 1998.

[80] H. Debar, B. Morin, F. Cuppens, F. Autrel, and L. Mé, “Détection d’intrusions: corrélation d’alertes,” *TSI. Tech. Sci. informatiques*, vol. 23, no. 3, pp. 359–390, 2004.

[81] “statistiques IOT.” https://informationmatters.net/internet-of-things-statistics/.

[82] F. Bonomi, R. Milito, P. Natarajan, and J. Zhu, “Fog computing: A platform for internet of things and analytics,” in *Big data and internet of things: A roadmap for smart environments*, Springer, 2014, pp. 169–186.