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 :

DJAAD MHAMED & DAHMANI CHERIF

Sur le thème

Vers Un Système De Détection D'intrusion Basée Sur Un Skyline De Classifieurs Dans Un Environnement Iots (L'internet Des Objets)

Soutenu publiquement à Tiaret devant le jury composé de :

Mr. DAHMANI Youcef	P.R	Université IBN-KHALDOUN Tiaret	Président
Mr. ALEM Abdelkader	M.A.A	Université IBN-KHALDOUN Tiaret	Encadreur
Mr. MOSTEFAOUI Sid Ahmed	M.C.A	Université IBN-KHALDOUN Tiaret	Examinateur

2021 – 2022

Remerciements

Nous tenons tout d'abord à remercier Allah le tout puissant etLe 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. ***DAHMANI Youcef*** *ET Mr.****MOSTEFAOUI Sid Ahmed***
D'avoir accepte d'examiner et d'évaluer notre travail.

Sans oublier ***Mr. BOUAZZA Abdelhamid*** *pour son soutien et son aide.*

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 ***2020-2022****.*

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.*
- *Ma petite famille, ma femme que je ne pourrai jamais. remercier assezpour son aide et son soutien ; mes enfants.*
- *A mes sœurs et frères et je les souhaite beaucoup de joie Etde bon heur dans la vie.*
- *A tout ma famille.*
- *A tous mes amis et mes camarades.*
- *A tous ceux qui me sont chers.*
- *A tous mes collègues de travail*

DJAAD MHAMED

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, mes enfants : Amina ET. Anes.*
- *Mes frères et sœurs.*
- *Tous mes amis et mes camarades.*
- *A tous ceux qui me sont chers.*

DAHMANI CHERIF

Résumé

L'**internet des Objets** (IdO) est l'évolution naturelle de l'utilisation des réseaux, il a pour objectif de rendre le monde réel plus intelligent grâce à la connexion des objets. Ces derniers obtiennent des informations qu'elles transmettent par le réseau. N'importe quel objet autonome qui peut être connecté à Internet et qui peut être utilisé à distance peut être considéré comme un membre de la famille de l'Internet des objets.

IOT a de nombreux protocoles de routage, le plus largement utilisé est le protocole RPL, qui prend compte tenu de la puissance limitée et des capacités de l'appareil, mais il souffre de plusieurs faiblesses, la plus importante est les attaques basées sur le routage qui ciblent ce protocole.

L'idée principal de ce travail est de proposer un IDS performant en détectant les attaques de protocole RPL dans un environnement IoT Nous avons proposé un modèle (hybride et hiérarchique) de deux niveaux, on a utilisé l'opérateur skyline pour sélectionner un seul niveau de façon automatique et un réseau bayésien naïf pour faire fusionner les différentes prédictions.

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 , Routage, Système de détection D'intrusion, Dodag, Protocol de Routage . Sécurité, Attaque…

Abstract

The **Internet of Things** (IoT) is the natural evolution of the use of networks , it aims to make the real world smarter through the connection of objects. The latter obtain information which they transmit through the network. Any autonomous object that can be connected to the Internet and can be used remotely can be considered a member of the Internet of Things family.

IOT has many routing protocols, the most widely used is the RPL protocol, which takes into account the limited power and capabilities of the device, but it suffers from several weaknesses, the most important is routing-based attacks that target this protocol.

The main idea of this work is to propose a powerful IDS by detecting RPL protocol attacks in an IoT environment We proposed a model (hybrid and hierarchical) of two levels, we used the skyline operator to select a single level automatically and a naive Bayesian network to merge the different predictions. The results obtained are very satisfactory because we manage to obtain a minimum rate of false alerts (false positive) and a maximum rate of attack detection, with a reduction in learning and prediction time.

Keywords: Internet Of Things , Routing , Intrusion detection System , Dodag ,Routing Protocol ,Security, Attack …

الملخص

إنترنت الأشياء (IoT) هو التطور الطبيعي لاستخدام الشبكات، فهو يهدف إلى جعل العالم الحقيقي أكثر ذكاءً من خلال توصيل الأشياء. يحصل هؤلاء الأخيرين على المعلومات التي ينقلونها عبر الشبكة. يمكن اعتبار أي كائن مستقل يمكن توصيله بالإنترنت ويمكن استخدامه عن بُعد عضوًا في عائلة إنترنت الأشياء.

يحتوي IOT على العديد من بروتوكولات التوجيه، وأكثرها استخدامًا هو بروتوكول RPL، والذي يأخذ في الاعتبار الطاقة والقدرات المحدودة للجهاز، لكنه يعاني من عدة نقاط ضعف، أهمها الهجمات القائمة على التوجيه التي تستهدف هذا البروتوكول.

الفكرة الرئيسية لهذا العمل هي اقتراح IDS قوي من خلال اكتشاف هجمات بروتوكول RPLفي بيئة إنترنت الأشياء. لقد اقترحنا نموذجًا (هجينًا وهرميًا) من مستويين ، استخدمنا مشغل الأفق لتحديد مستوى واحد تلقائيًا وساذج شبكة بايز لدمج التوقعات المختلفة .النتائج التي تم الحصول عليها مُرضية للغاية لأننا نجحنا في الحصول على معدل أدنى من التنبيهات الخاطئة (إيجابية كاذبة) ومعدل أقصى لاكتشاف الهجوم ، مع تقليل وقت التعلم والتنبؤ .

الكلمات الرئيسية : انترنت الاشياء ،توجيه ، نظام كشف التسلل ، بروتوكول التوجيه منخفض الطاقة، الأمان، هجوم ...

Table des Matières

Chapitre 02 : Internet des Objet IoT

Chapitre 3 : Techniques de classification et Calcul de Skyline

Chapitre 4 : Contribution dans la Détection D'intrusion

Chapitre 5 : Réalisation et implémentation

LISTE DES FIGURES :

Liste des Tableaux :

Liste des figure

Introduction Générale

INTRODUCTION GENERALE

L'IoT en tant qu'une évolution de l'internet actuelle et étant un réseau mondial d'objets interconnectés, elle permet une amélioration considérable de notre mode de vie et la façon dont les objets intelligents dans notre entourage interagissent entre eux et avec leurs utilisateurs.

Ces objets, qui sont considérés comme la plateforme de base de l'IoT basés sur l'utilisation de différents protocoles de communication. La sécurité de l'IoT reste encore un des problèmes majeurs qui freine l'évolution et le déploiement rapide de cette technologie dû à la multiplication des objets connectés qui deviennent des cibles pour les pirates.

La détection des tentatives d'attaques sur un réseau est une problématique très importante dans le domaine de l'IoT. Les technologies classiques de protection des réseaux de type Firewall sont en effet inefficaces contre la plupart des attaques actuelles. . Aussi sont apparus de nouveaux équipements réseaux pour prendre en compte ces carences, les systèmes de détection d'intrusions (IDS).

Le principal rôle clé d'un système de détection d'intrusion (IDS) est de détecter les actions qui essaient de compromettre la confidentialité, l'intégrité ou la disponibilité d'une ressource.

Cependant, à son tour, ils présentent également certaines faiblesses. En effet, certaines attaques peuvent passer inaperçues (faux négatifs), ou encore certaines alertes peuvent être générées par rapport à des attaques qui n'ont pas eu lieu (faux positif). En outre, le nombre d'alertes générées est souvent trop élevé si bien que l'opérateur de sécurité qui est chargée d'analyser et de traiter ces alertes se retrouve rapidement noyé.

Dans ce contexte, Notre travail consiste à proposer un nouveau modèle d'un IDS comportementale qui soit efficace, simple à mettre en œuvre et qui ne nécessitent pas de connaissances d'experts. Il doit non seulement être capable de minimiser le taux de fausses alertes mais aussi trouver un point d'équilibre entre les deux autres mesures de performance (taux d'exactitude, taux de détection) tout en maximisant le taux de détection des attaques rares.

L'idée principale est d'utiliser l'opérateur skyline pour choisir les meilleurs classifieurs à hybrider et un réseau bayésien naïf pour faire fusionner les différentes prédictions.

Avant de présenter l'approche proposée au cours de notre travail, nous allons tout d'abord commencer par parler, dans le premier chapitre, de la sécurité des réseaux informatiques, des attaques réseaux courantes, des mécanismes de sécurité, et dans le deuxième chapitre, les systèmes de détection d'intrusion et les approches de détection. Dans le troisième chapitre, nous allons détailler quelques techniques d'apprentissage et le principe du calcul de skyline dans un environnement décisionnel. Dans le quatrième chapitre, nous présenterons l'approche proposée. En fin Nous présentons l'ensemble des outils de développement utilisés.

Chapitre 1

La sécurité informatique et IDS

Chapitre 1 : Sécurité des Réseaux informatiques et IDS.

Introduction :

L'informatique et en particulier l'Internet jouent un rôle grandissant dans le domaine des réseaux. Un grand nombre d'applications critiques d'un point de vue de leur sécurité sont déployées dans tous les secteurs professionnels. La sécurité de ces réseaux constitue un enjeu crucial , le contrôle des informations traitées et partagées au sein de ces réseaux devient alors un problème majeur d'autant plus que les réseaux sont interconnectés entre eux, ce qui complexifie donc la tâche des responsables de la sécurité.

Tout au long de ce chapitre, notre intérêt se portera sur les principales menaces pesant sur la sécurité des réseaux ainsi que les mécanismes de défense.

1 Définition Sécurité des réseaux :

La sécurité d'un réseau est un ensemble de moyens techniques, organisationnels, juridiques et humains nécessaires et mis en place pour conserver, rétablir, et garantir sa sécurité contre les menaces accidentelles ou intentionnelles. En général, la sécurité d'un réseau englobe celle du système informatique sur lequel il s'appuie. [1]

2 Les critères de la sécurité d'un réseau :

Les objectifs d'une politique de sécurité sont de garantir la sécurité des informations et des réseaux d'entreprise. Ces impératifs peuvent être définis à plusieurs niveaux : [1]

2.1 L'intégrité :

Consiste à déterminer si les données n'ont pas été altérées durant la communication. C'est-à-dire garantir que les données sont bien celles que l'on croit être.

2.2 La confidentialité :

Consistant à assurer que seules les personnes autorisées aient accès aux ressources échangées.

2.3 La disponibilité :

L'objectif de la disponibilité est de garantir l'accès à un service ou à des ressources, permettant de maintenir le bon fonctionnement du système d'information quand les informations sont accessibles au moment voulu.

2.4 Non répudiation :

Permettant de garantir qu'une transaction ne peut être niée . La non-répudiation de l'origine prouve que les données ont été envoyées , et la non-répudiation de l'arrivée prouve qu'elles ont été reçues.

2.5 L'authentification :

Consistant à assurer que seules les personnes autorisées aient accès aux ressources. Elle consiste à assurer l'identité d'un utilisateur, c'est-à-dire de garantir à chacun des correspondants que son partenaire est bien celui qu'il croit être. Un contrôle d'accès peut permettre l'accès à des ressources uniquement aux personnesautorisées [1].

3 Les Causes pour sécuriser les réseaux :

3.1 Les enjeux :

3.1.1 Enjeux économique :

Les organismes ou entreprises à but lucratif ont presque toujours la même finalité : c'est de réaliser des bénéfices sur l'ensemble de leurs activités. Cette réalisation est rendue possible grâce à son système d'information considéré comme moteur de développement de l'entreprise. D'où la nécessité de garantir la sécurité de ce dernier. La concurrence fait que des entreprises s'investissent de plus en plus dans la sécurisation de leurs systèmes d'information et dans la qualité de service fournit aux clients.

3.1.2 Enjeux politiques :

La plupart des entreprises ou organisations se réfèrent aux documents officiels de sécurité élaborés et recommandés par l'État. Ces documents contiennent généralement des directives qui doivent être appliquées par toute structure engagée dans un processus de sécurisation du réseau. Dans le cadre du chiffrement des données par exemple, chaque État définit des cadres et mesures d'utilisation des algorithmes de chiffrement et les recommande aux entreprises exerçant sur son territoire. Le non-respect de ces mesures et recommandations peut avoir des conséquences graves sur l'entreprise. A ce niveau, l'enjeu est plus politique parce que chaque État souhaite être capable de décrypter toutes les informations circulant dans son espace [1].

3.1.3 Enjeux juridiques :

Dans un réseau, on retrouve de l'information multiforme (numérique, papier, etc.). Le traitement de celle-ci doit se faire dans un cadre bien définit et dans le strict respect des lois en vigueur. En matière de juridiction, le non- respect des lois et exigences relatives à la manipulation des informations dans un système d'information peut avoir des conséquences graves sur l'entreprise [1].

3.2 Les Vulnérabilités :

Tous les systèmes informatiques sont vulnérables. Peu importe le niveau de vulnérabilité de ceux-ci. Une vulnérabilité est une faille ou une faiblesse pouvant être exploitée par une personne mal intentionnée pour nuire.

Les vulnérabilités des systèmes peuvent être classées en catégorie (humaine, technologique, organisationnelle, mise en œuvre) [1].

3.2.1 Vulnérabilités humaines :

L'être humain de par sa nature est vulnérable. La plupart des vulnérabilités humaines proviennent des erreurs (négligence, manque de compétences, surexploitation, etc.), car ne dit-on pas souvent que l'erreur est humaine? Un SI étant composé des humains, il convient d'assurer leur sécurité si l'on veut garantir un maximum de sécurité dans le SI .

3.2.2 Vulnérabilités technologiques :

Avec la progression exponentielle des outils informatiques, les vulnérabilités technologiques sont découvertes tous les jours. Ces vulnérabilités sont à la base dues à une négligence humaine lors de la conception et la réalisation. Pour être informé régulièrement des vulnérabilités technologiques découvertes, il suffit de s'inscrire sur une liste ou des listes de diffusion mises en Place par les CERT (Computer Emergency Readiness ou Response Team) .

3.2.3 Vulnérabilités organisationnelles :

Les vulnérabilités d'ordre organisationnel sont dues à l'absence des documents cadres et formels, des procédures (de travail, de validation) suffisamment détaillées pour faire face aux problèmes de sécurité du système. Quand bien même ces documents et procédures existent, leur vérification et mises à jour ne sont pas toujours bien assurées .

3.2.4 Vulnérabilités mise en œuvre :

Les vulnérabilités au niveau mise en œuvre peuvent être dues au non prise en compte de certains aspects lors de la réalisation d'un projet.

3.3 Les Menaces :

Une menace est un événement, d'origine accidentelle ou délibérée, capable s'il se réalise de causer un dommage au sujet étudié. Le réseau informatique comme tout autre réseau informatique est en proie à des menaces de toutes sortes qu'il convient de recenser.

On peut également classer les menaces en deux catégories :

3.3.1 Les menaces passives :

Consistent essentiellement à copier ou à écouter l'information sur le réseau, elles nuisent à la confidentialité des données. Dans ce cas, celui qui prélève une copie n'altère pas l'information elle-même.

3.3.2 Les menaces actives :

Consistent à altérer des informations ou le bon fonctionnement d'un service.

3.4 Les logiciels malveillants :

Ce sont des logiciels développés par des hackers dans le but de nuire à un système d'informations.

3.4.1 Les Virus :

Un virus est un segment de programme qui, lorsqu'il s'exécute, se reproduit en s'adjoignant à un autre programme (du système ou d'une application), et qui devient ainsi un cheval de Troie. Puis le virus peut ensuite se propager à d'autres ordinateurs (via un réseau) à l'aide du programme légitime sur lequel il s'est greffé. Il peut également avoir comme effets de nuire en perturbant plus ou moins gravement le fonctionnement de l'ordinateur infecté [2].

Les virus peuvent être classés suivant leur mode de propagation et leurs cibles : [3]

3.4.1.1 Le virus de boot :

il est chargé en mémoire au démarrage et prend le contrôlede l'ordinateur.

3.4.1.2 Le virus d'application :

Ils infectent les programmes exécutables, c'est-à-dire les programmes (.exe, .com ou .sys) en remplaçant l'amorce du fichier, de manière à ce que le virus soit exécuté avant le programme infecté. Puis ces virus rendent la main au programme initial, camouflant ainsi leur exécution aux yeux de l'utilisateur [2].

3.4.1.3 La macro virus :

Il infecte des logiciels de la suite Microsoft Office les documents bureautiques en utilisant leur langage de programmation, qui contaminera tous les documents basés sur lui, lors de leur ouverture.

3.4.2 Les Vers :

Un ver est un programme autonome qui se reproduit et se propage à l'insu des utilisateurs. Contrairement aux virus, un ver n'a pas besoin d'un logiciel hôte pour se dupliquer. Le ver a habituellement un objectif malicieux, par exemple

- espionner l'ordinateur dans lequel il réside ;
- Offrir une porte dérobée à des pirates informatiques.
- Détruire des données sur l'ordinateur infecté.
- envoyer de multiples requêtes vers un serveur internet dans le but de le saturer.

3.4.3 Les chevaux de Troie :

Un cheval de Troie est une forme de logiciel malveillant déguisé en logiciel utile. Son but : se faire exécuter par l'utilisateur, ce qui lui permet de contrôler l'ordinateur et de s'en servir pour ses propres fins. Généralement d'autres logiciels malveillants seront installés sur votre ordinateur, tels que permettre la collecte frauduleuse, la falsification ou la destruction de données.

3.4.4 Les logiciels espions :

Est un programme ou un sous- programme, conçu dans le but de collecter des données personnelles sur ses utilisateurs et de les envoyer à son concepteur, ou à un tiers via Internet ou tout autre réseau informatique, sans avoir obtenu au préalable une autorisation explicite et éclairée desdits utilisateurs.

3.4.5 Le spam :

Correspond à l'envoi intempestif de courriers électroniques, publicitaires ou non, vers une adresse mail. Le spam est une pollution du courrier légitime par une énorme masse de courrier indésirable non sollicité [4].

3.5 Les intrusions :

Une intrusion est définie comme une faute malveillante d'origine interne ou externe résultant d'une attaque qui a réussi à exploiter une vulnérabilité. Elle est susceptible de produire des erreurs pouvant provoquer une défaillance vis-à-vis la sécurité, c'est-à-dire une violation de la politique de sécurité du système. Le terme d'intrusions sera employé dans le cas où l'attaque est menée avec succès et où l'attaquant a réussi à s'introduire et/ou compromettre le système [5].

3.6 Les Attaques :

3.6.1 Définition :

Une attaque est définie comme faute d'interaction malveillante visant à violer une ou plusieurs propriétés de sécurité. C'est une faute externe créée avec l'intention denuire, y compris les attaques lancées par des outils automatiques : vers, virus, etc. La notion d'attaque ne doit pas être confondue avec la notion d'intrusions [5].

3.6.2 Les motivations d'une attaque :

Les motivations des attaques peuvent être liées à divers objectif [6]:

- Obtenir un accès au système.
- Voler des informations, tels que des secrets industriels ou des propriétésintellectuelles.
- Collectionner des informations personnelles sur un utilisateur
- S'informer sur l'organisation.

- Récupérer des données bancaires.
- 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, notamment lorsque le réseau sur lequel il est situé possède une bande passante élevée.
- Faire du chantage.
- Par simple jeu ou par défi.
- Pour terrorisme ou pour des fins politique.
- Pour apprendre.

3.6.3 Type d'attaques :

Les hackers utilisent plusieurs techniques d'attaques. Ces attaquespeuvent être regroupées en trois familles différentes [7]:

3.6.3.1 Les attaques directes :

C'est la plus simple des attaques. Le hacker attaque directement sa victime à partir de son ordinateur. La plus part des hackers utilisent cette technique. En effet, les programmes de hack qu'ils utilisent ne sont que faiblement paramétrable, et un grand nombre de ces logiciels envoient directement les paquets à la victime.

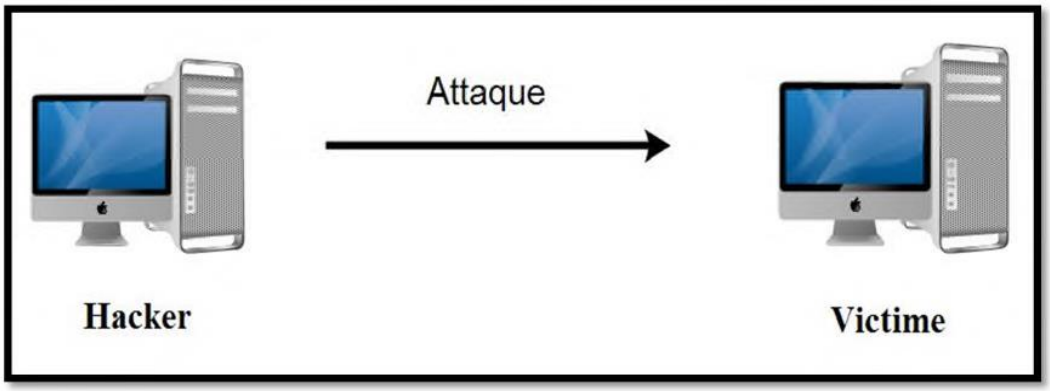

Figure 1.1 : Attaque directe

3.6.3.2 Les attaques indirectes par rebond :

Cette attaque est très prisée des hackers.En effet, le rebond a deux avantages :

Masquer l'identité du hacker.

Utiliser les ressources de l'ordinateur intermédiaire car il est plus puissantpour attaquer.

Le principe en lui-même, est simple : Les paquets d'attaque sont envoyés à l'ordinateur intermédiaire, qui répercute l'attaque vers la victime. D'où le terme par rebond.

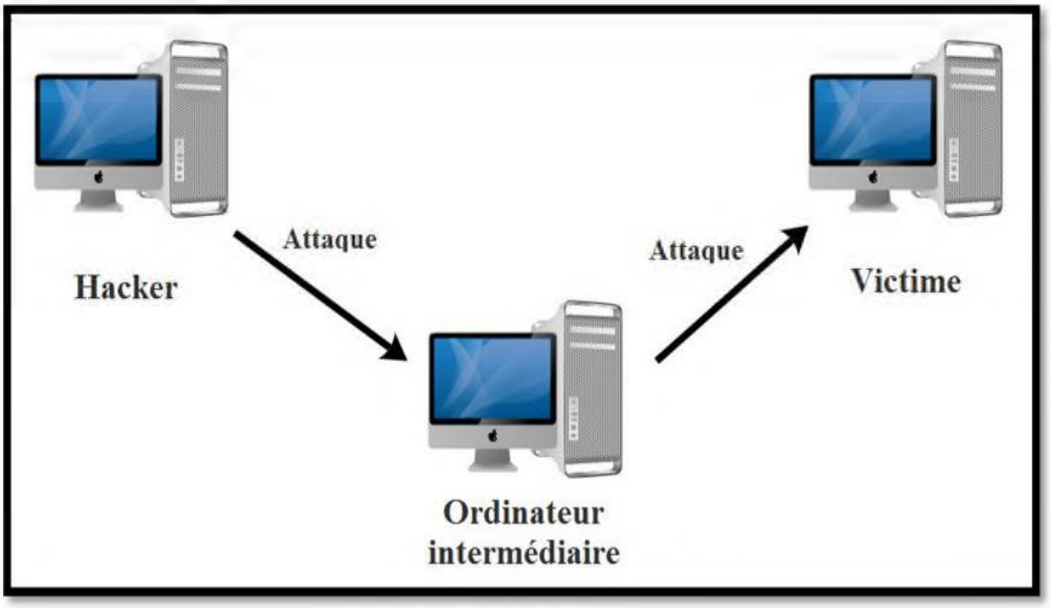

Figure 1.2 : Attaque indirecte par rebond

3.6.3.3 Les attaques indirectes par réponse :

Cette attaque est un dérivé de l'attaque parrebond. Elle offre les mêmes avantages, du point de vue du hacker. Mais au lieu d'envoyer une attaque à l'ordinateur intermédiaire pour qu'il la répercute, l'attaquant va lui envoyer une requête. Et c'est cette réponse à la requête qui va être envoyée à l'ordinatrice victime.

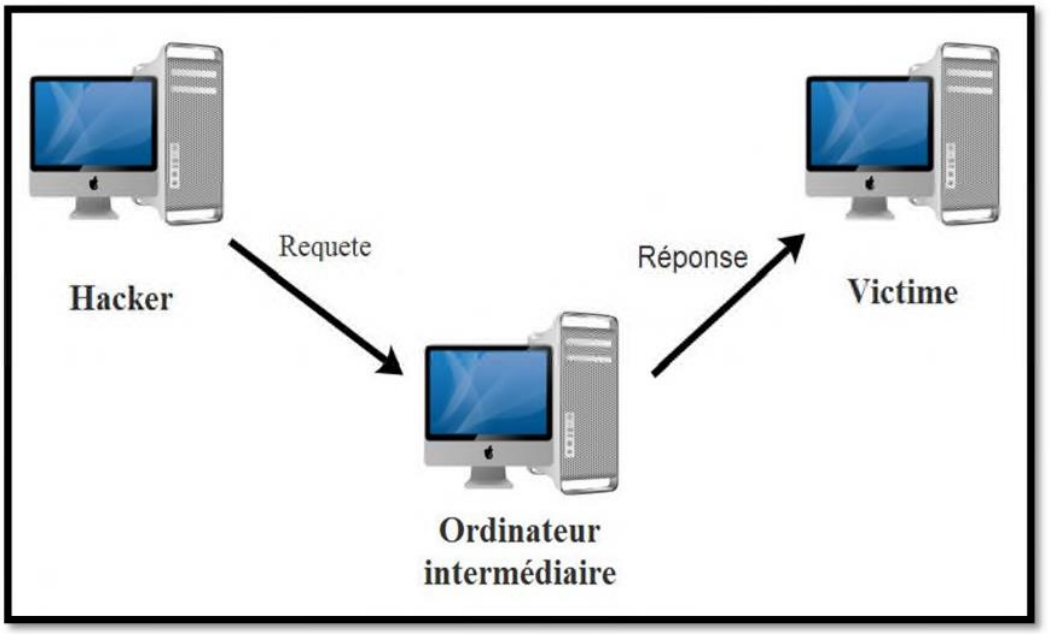

Figure 1.3 : Attaque indirecte par réponse

3.7 Les moyens de sécurisé un réseau :

La sécurité d'un réseau c'est la sécurité des éléments qui le compose, il existe plusieurs mécanismes et dispositifs de sécurité, parmi eux :

3.7.1 Les Antivirus :

Les antivirus sont des logiciels conçus pour identifier, neutraliser et éliminer des logiciels malveillants. Ceux-ci peuvent se baser sur l'exploitation de failles de sécurité, mais il peut également s'agir de programmes modifiant ou supprimant des fichiers, que ce soit des documents de l'utilisateur de l'ordinateur infecté, ou des fichiers nécessaires au bon fonctionnement de l'ordinateur.

Un antivirus vérifie les fichiers et courriers électroniques, les secteurs de boot (pour détecter les virus de boot), mais aussi la mémoire vive de l'ordinateur, les médias amovibles (clefs USB, CD, DVD, etc.), les données qui transitent sur les éventuels réseaux (dontinternet), etc.

3.7.2 Les mises à jour système :

Pour éviter les dénis de services applicatifs, on doit maintenir tous les logiciels de son système à jour puisque les mises à jour permettent souvent de corriger des failles logicielles, qui peuvent être utilisées par un attaquant, pour mettre l'application hors service, ou pire, le serveur. Il est donc impératif de mettre son système à jour très régulièrement C'est un moyen très simple à mettre en place pour se protéger des attaques applicative.

Editer des options dans les fichiers de configuration qui stocke des données concernant chaque connexion reçue par la machine telle l'adresse IP source, le numéro de port, l'âge dela connexion. En analysant ces données, on peut facilement détecter les comportementssuspects et éviter certains types d'attaque.

3.7.3 Les firewalls :

En français on dit pare-feu ou garde-barrière, c'est un système permettant de protéger un ordinateur ou un réseau d'ordinateurs des intrusions provenant d'un réseau tiers (notamment internet). Le pare-feu est un système permettant de filtrer les paquets de données échangés avec le réseau, il s'agit ainsi d'une passerelle filtrante comportant au minimum les interfaces réseau suivante :

- une interface pour le réseau à protéger (réseau interne) .
- une interface pour le réseau externe.

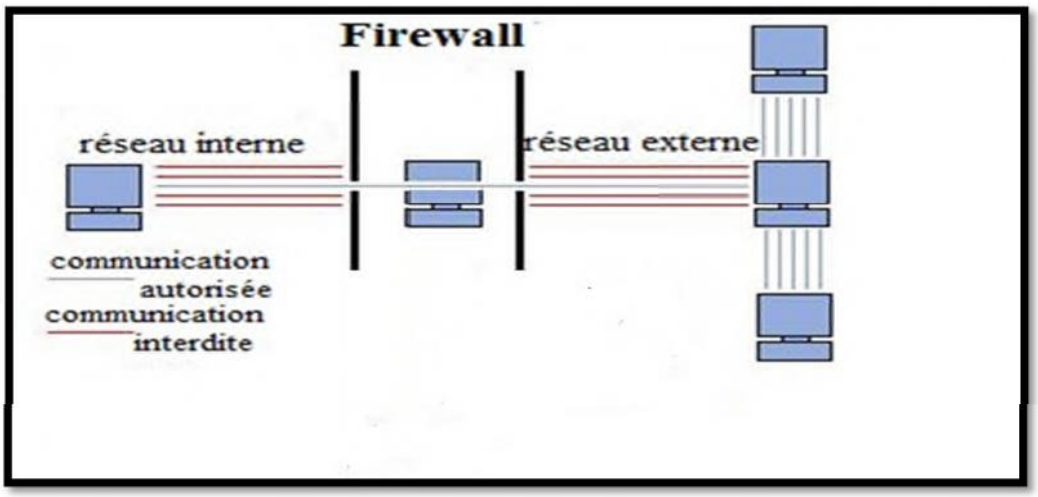

Figure 1.4 : Firewall

Le système firewall est un système logiciel, reposant parfois sur un matériel réseau dédié, constituant un intermédiaire entre le réseau local (ou la machine locale) et un ou plusieurs réseaux externes. Il est possible de mettre un système pare-feu sur n'importe quelle machine et avec n'importe quel système pourvu que :

- La machine soit suffisamment puissante pour traiter le trafic.
- Le système soit sécurisé.

Aucun autre service que le service de filtrage de paquets ne fonctionne sur leserveur.

3.7.4 Fonctionnement d'un système pare-feu :

C'est un ensemble de différents composants matériels (physique) et logiciels (logique)qui contrôlent le trafic intérieur/extérieur selon une politique de sécurité. Un système pare-feu fonctionne la plupart du temps grâce à des règles de filtrage indiquant les adresses IP autorisées à communiquer avec les machines aux réseaux, il s'agit ainsi d'une passerelle filtrante.

Il permet d'une part de bloquer des attaques ou connexions suspectes d'accéder au réseau interne.

D'un autre côté, un firewall sert dans de nombreux cas également à éviter la fuite non contrôlée d'informations vers l'extérieur. Il propose un véritable contrôle sur le trafic réseau de l'entreprise, Il permet donc d'analyser, de sécuriser et de gérer le trafic réseau.

3.7.5 Architecture DMZ :

Une DMZ est une zone d'un réseau d'entreprise, située entre le réseau local et Internet, derrière le pare-feu. Il s'agit d'un réseau intermédiaire regroupant des serveurs ou services (HTTP, DHCP, mails, DNS, etc.). Ces serveurs devront être accessibles depuis le réseau interne de l'entreprise et, pour certains, depuis les réseaux externes. Le but est ainsi d'éviter toute connexion directe au réseau interne.

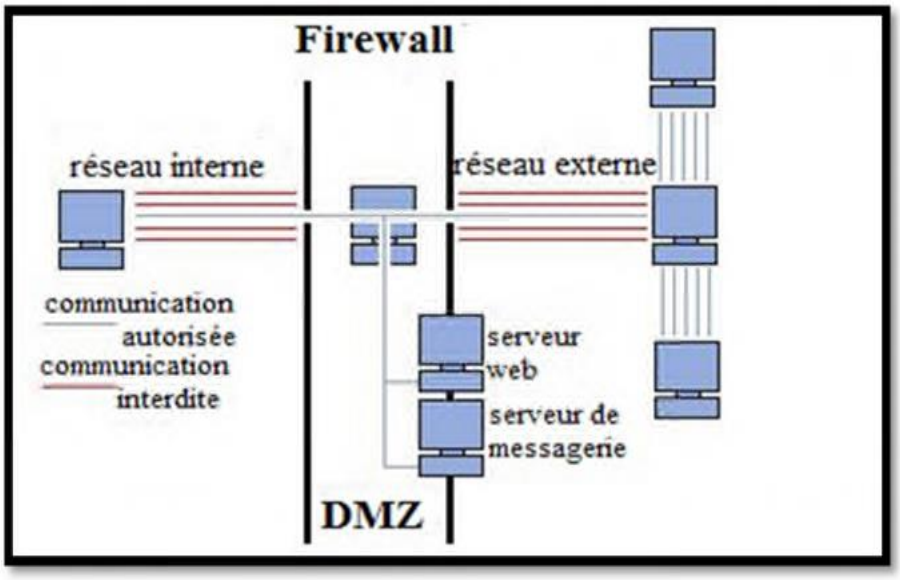

Figure 1.5 : Architecture DMZ

3.7.6 Les VPN: (Virtual Private network)

Dans les réseaux informatiques, le réseau privé virtuel est une technique permettant aux postes distants de communiquer de manière sûre, tout en empruntant des infrastructures publiques (internet).

Un VPN repose sur un protocole, appelé protocole de tunnelisation, c'est-à-dire un protocole permettant aux données passant d'une extrémité à l'autre du VPN d'être sécurisées par des algorithmes de cryptographie [8].

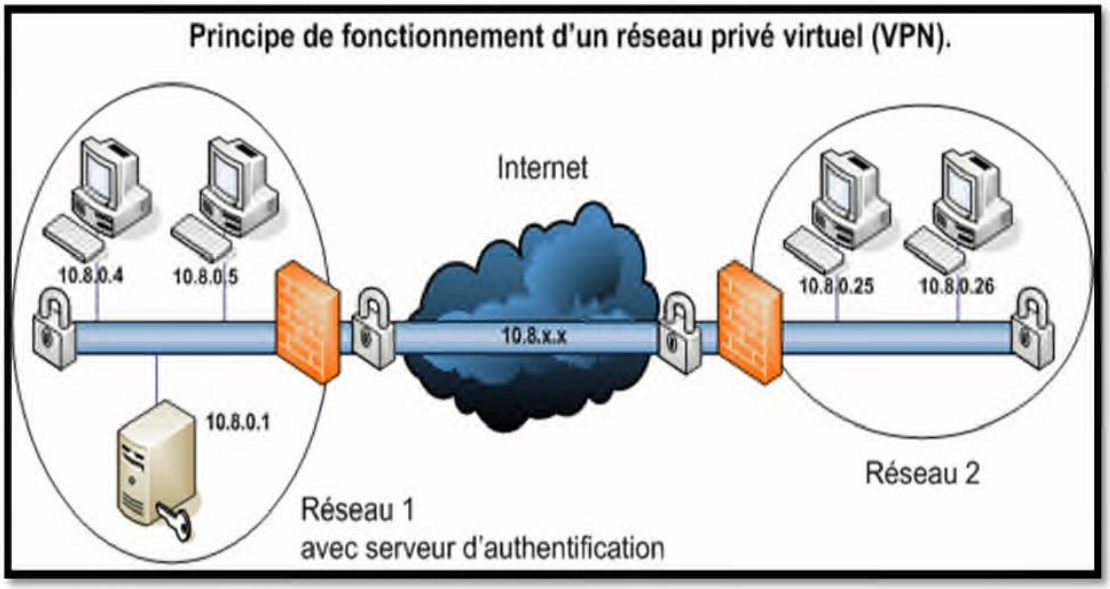

Figure 1.6 : Principe d'un VPN

3.7.7 Les Système de détection d'intrusion (IDS) :

La détection d'intrusion est définie comme étant un mécanisme écoutant le trafic réseau de manière furtive, afin de repérer des activités anormales ou suspectes et permettant ainsi d'avoir une stratégie de prévention sur les risques d'attaques. Il existe différents types d'IDS(Voir la partie d'IDS) :

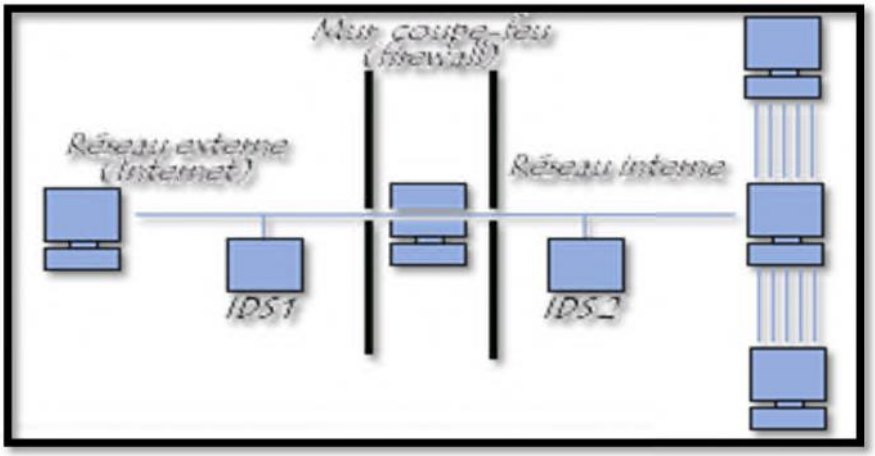

Figure 1.7: Système de détection d'intrusions

3.7.8 Les Algorithme de chiffrements :

Il existe deux grandes familles d'algorithmes de chiffrements, ceux à clés symétriques et ceux à clés asymétriques.

- **Algorithme de chiffrement symétrique** : Il consiste à utiliser la même clé pour le chiffrement ainsi que pour le déchiffrement. Il est donc nécessaire que les deux interlocuteurs se soient mis d'accord sur une clé privée, ou ils doivent utiliser un canal sécurisé pour l'échanger.
- **Algorithme de chiffrement asymétrique :** C'est une méthode cryptographique faisant intervenir une paire de clés asymétrique (une clé publique et une clé privée). Elle utilise cette paire de clés pour le chiffrement et le déchiffrement. La clé publique est rendue publique et elle est distribuée librement, la clé privée quant à elle n'est jamais distribuée et doit être gardé secrète.

4 Mise en œuvre d'une politique de sécurité :

La politique de sécurité des systèmes d'information est un plan d'actions définies pour maintenir un certain niveau de sécurité. Elle reflète la vision stratégique de la direction de l'entreprise en matière de sécurité des systèmes d'informations (SSI).

Une politique de sécurité s'élabore à plusieurs niveaux :

Sécuriser l'accès aux données de façon logicielle (authentification, contrôle d'intégrité).

Sécuriser l'accès physique aux données : serveurs placés dans des salles blindées avec badge d'accès…

Un aspect très important pour assurer la sécurité des données d'une entreprise est de sensibiliser les utilisateurs aux notions de sécurité, de façon à limiter les comportements à risque : si tout le monde peut accéder aux salles de serveurs, peut imposte qu'elles soient sécurisées !

De même, si les utilisateurs laissent leurs mots de passes écrit à côté de leur PC, son utilité est limitée…

Enfin , il est essentiel pour un responsable de sécurité de s'informer continuellement, des nouvelles attaques existantes, des outils disponibles…de façon à pouvoir maintenir à jour son système de sécurité et à combler les brèches de sécurité qui pourraient exister.

***Partie 2 : Les systèmes de détection d'intrusions* :**

Introduction :

En sécurité informatique, la détection d'intrusion est l'acte de détecter les actions qui essaient de compromettre la confidentialité, l'intégrité ou la disponibilité d'une ressource.

Les systèmes de détection d'intrusions (IDS) sont généralement considérés comme une seconde ligne de défense pour protéger contre les activités malicieuses.

La détection des tentatives d'attaques sur un réseau est une problématique très importante dans le domaine de la sécurité informatique. Les technologies classiques de protection des réseaux de type Firewall filtrant sont en effet inefficaces contre la plupart des attaques actuelles. Aussi sont apparus de nouveaux équipements réseaux pour prendre en compte ces carences, les NIDS, systèmes de détection d'intrusions réseaux, dont le but est de détecter les tentatives d'attaque qu'un firewall ne peut pas bloquer. Malheureusement, en pratique, les NIDS génèrent tellement d'alertes sur un réseau important qu'il en devient très difficile de déterminer celles générés par une attaque réelle. Le présent chapitre détaille les IDS ainsi que les différentes approches de détections qui existante.

5 Définition d'un système d'intrusion :

IDS signifie Intrusion Détection System. Il s'agit d'un équipement permettant de surveiller l'activité d'un réseau ou d'un hôte donné, afin de détecter toute tentative d'intrusion et éventuellement de réagir à cette tentative.

IDS est un appareil ou une application qui alerte l'administrateur en cas de faille de sécurité, de violation de règles ou d'autres problèmes susceptibles de compromettre son réseau informatique.

Un IDS (Intrusion Détection System) est un ensemble de composants logiciels et/ou matériels dont la fonction principale est de détecter et analyser des activités anormales ou suspectes sur la cible analysée (un réseau ou un hôte). Il permet ainsi d'avoir une connaissance sur les tentatives réussies comme échouées des intrusions.

Certains termes sont souvent employés quand on parle d'IDS :

- **Faux positif**:une alerte provenant d'un IDS, mais qui ne correspond pas à une attaque réelle.
- **Faux négatif** : une intrusion réelle qui n'a pas été détectée par l'IDS.

6 Les types des systèmes de détection d'intrusion :

A cause de la diversité des attaques que mettent en œuvre les pirates, la détection d'intrusion doit se faire à plusieurs niveaux. Il existe donc différents types d'IDS :

6.1 Les NIDS (Network-based Intrusion Detection System):

Les NIDS sont des IDS dédiés aux réseaux. Ils comportent généralement une machine qui écoute sur le segment de réseau à surveiller, un capteur et un moteur qui réalise l'analyse du trafic afin de détecter les intrusions en temps réel. Un NIDS écoute donc tout le trafic réseau, puis l'analyse et génère des alertes si des paquets semblent dangereux [10].

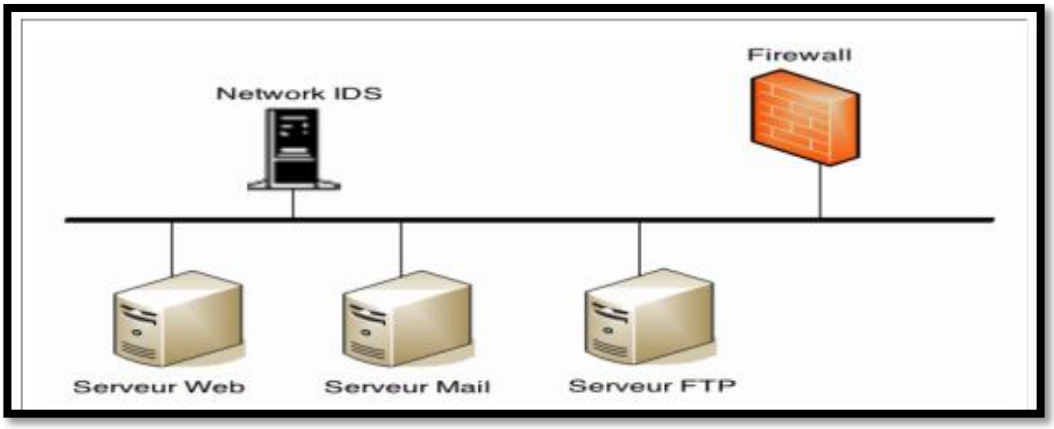

Figure 1.8 système de détection d'intrusion réseau

L'implantation d'un NIDS sur un réseau se fait de la façon suivante: des capteurs (souvent de simples hôtes) sont placés aux endroits stratégiques du réseau et génèrent les alertes s'ils détectent une attaque. Ces alertes sont envoyées à une console sécurisé qui les analyse et les traite éventuellement. Cette console est généralement située sur un réseau isolé qui relie uniquement les capteurs et la console [9].

On peut placer les capteurs dans deux endroits différents :

- **A l'intérieur du pare-feu** : Si les capteurs se trouvent à l'intérieur du pare-feu, il sera plus facile de dire si le pare-feu a été mal configuré et nous pouvons ainsi savoir si une attaque est venue par ce pare-feu.
- **A l'extérieur du pare-feu** : Les capteurs placés à l'extérieur du pare-feu servent à la détection et l'analyse d'attaques. Il offre l'avantage d'écrire dans les logs, ainsi l'administrateur voit ce qu'il doit modifier dans la configuration du pare-feu.

6.2 Les HIDS (Host-based Intrusion Detection System):

Les systèmes de détection d'intrusion basés sur l'hôte analysent exclusivement l'information concernant cet hôte. Comme ils n'ont pas à contrôler le trafic du réseau mais seulement les activités d'un hôte ils se montrent habituellement plus précis sur les types d'attaques. De plus, nous remarquons immédiatement l'impact sur la machine concernée comme par exemple si un utilisateur l'attaquait avec succès. Ces IDS utilisent deux types de sources pour fournir une information sur l'activité : les logs et les traces d'audit du système d'exploitation. Chacun a ses avantages : les traces d'audit sont plus précises et détaillées et fournissent une meilleure information alors que les logs qui ne fournissent que l'information essentielle sont plus petits [12].

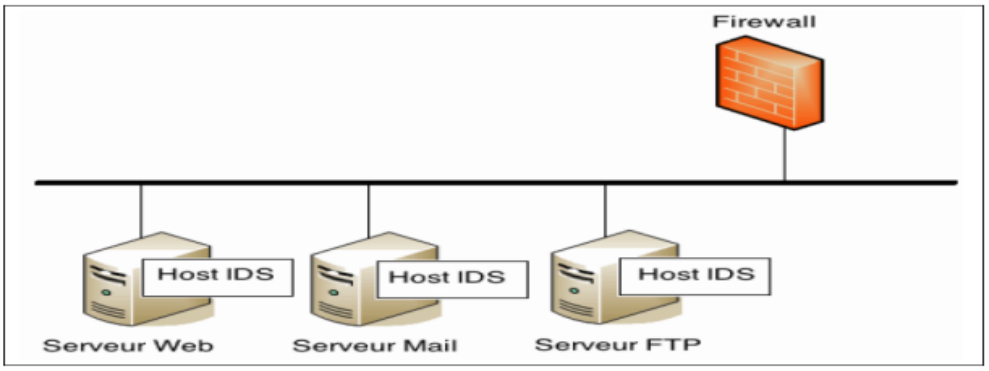

Figure 1.9 système de détection d'intrusion d'hôte

6.3 IDS hybride :

Les systèmes de détections d'intrusions hybrides, rassemblent les caractéristiques de plusieurs systèmes de détections différents. En pratique, on trouve la combinaison des NIDS et HIDS qui permettent de surveiller le réseau et l'hôte. Les sondes agissent comme un NIDS ou un HIDS Il permet de réunir les informations de diverses sondes placées sur le réseau. L'exemple le plus connu dans le monde Open-Source est Prélude.

Cet IDS permet de stocker dans une base de données des alertes provenant de différents systèmes relativement variés.

Utilisant Snort comme NIDS, et d'autres logiciels tels que Samhain en tant que HIDS,il permet de combiner des outils puissants tous ensemble pour permettre une visualisation centralisée des attaques .

7 Caractéristiques d'un système de détection d'intrusion :

Parmi les caractéristiques souhaitables trouvées dans un système de détection d'intrusion nous pouvons citer :

- Résister aux tentatives de corruption, c'est-à-dire, il doit pouvoir détecter s'il a subi lui-même une modification indésirable.

- Utiliser un minimum de ressources de système sous surveillance. S'adapter au cours du temps aux changements du système surveillé et du comportement des utilisateurs.
- Etre Utilisation minimale de ressources (de calcul, de stockage, etc.) sur le système sur lequel il est installé.
- Capacité d'accepter des mises à jour et des modifications de configuration pour rendre compte des nouvelles dispositions de la politique de sécurité et les changements susceptibles de s'opérer dans l'organisation (nouvelles acquisitions, restructuration,...etc.).
- Facilite et simplicité de déploiement : facilite d'installation et de configuration Portabilité,...etc.
- facilement configurable pour implémenter une politique de sécurité spécifique d'un réseau.

8 L'architecture d'un IDS :

Cette section décrit les trois composants qui constituent classiquement un système de détection d'intrusions. La Figure 1.15 illustre les interactions entre ces trois composants [6].

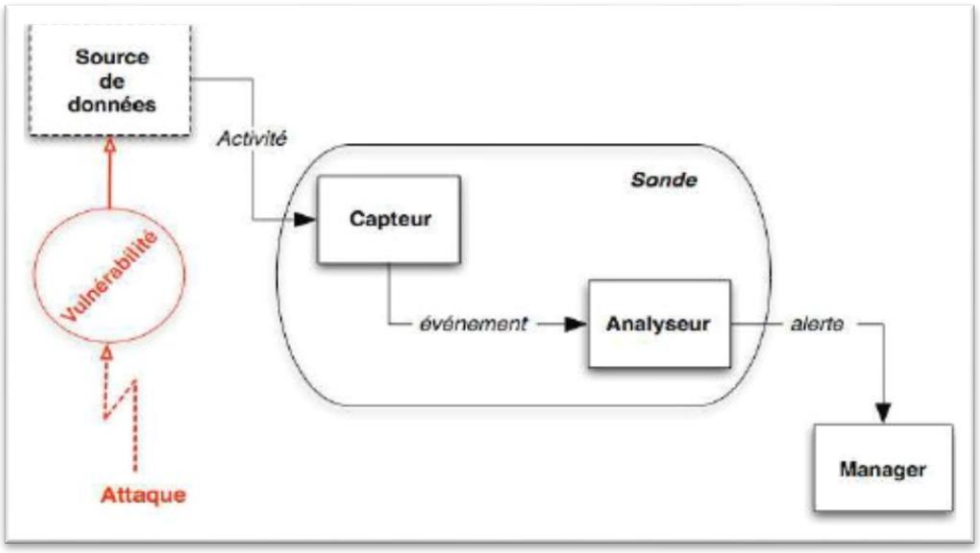

Figure 1.15 : Architecteur d'un IDS

8.1 Capteur :

Génère des événements en filtrant et formatant les données brutes provenant d'une source de données.

8.2 Analyseur :

L'objectif de l'analyseur est de déterminer si le flux d'événements fourni par le capteur contient des éléments caractéristiques d'une activité malveillante.

8.3 Manager :

Le manager collecte les alertes produites par le capteur, les met en forme et les présente à l'opérateur. Éventuellement, le manager est chargé de la réaction à adopter qui peut être :

- Confinement de l'attaque, qui a pour but de limiter les effets de l'attaque.
- Eradication de l'attaque, qui tente d'arrêter l'attaque.
- Recouvrement, qui est l'étape de restauration du système dans un état sain.
- Diagnostic, qui est la phase d'identification du problème.

9 Mise en place d'un IDS :

Le positionnement de l'IDS : Il existe plusieurs endroits stratégiques où il convient de placer un IDS. Le schéma suivant illustre un réseau local ainsi que les trois positions que peut y prendre un IDS [13]:

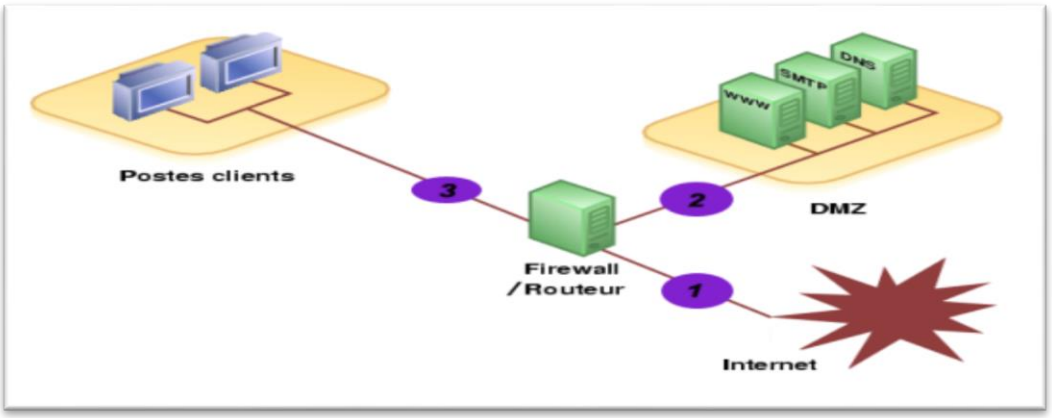

Figure1.16: la position IDS

- **Position (1):**

Sur cette position, l'IDS va pouvoir détecter l'ensemble des attaques frontales, provenant de l'extérieur, en amont du firewall. Ainsi, beaucoup d'alertes seront remontées ce qui rendra les logs difficilement consultables.

- **Position (2):**

Si l'IDS est placé sur la DMZ, il détectera les attaques qui n'ont pas été filtrées par le firewall et qui relèvent d'un certain niveau de compétence. Les logs seront ici plus clairs à consulter puisque les attaques ne seront pas recensées.

- **Position (3):**

L'IDS peut ici rendre compte des attaques internes, provenant du réseau local de l'entreprise. Il peut être judicieux d'en placer un à cet endroit étant donné le fait que 80% des attaques proviennent de l'intérieur. De plus, si des trojans ont contaminé le parc informatique (navigation peu méfiante sur internet) ils pourront être ici facilement identifiés pour être ensuite éradiques.

10 Mode de fonctionnement :

Il faut distinguer deux aspects dans le fonctionnement d'un IDS : le mode de détection utilisé et la réponse apportée par l'IDS lors de la détection d'une intrusion.

Il existe deux modes de détection, la détection d'anomalies et la reconnaissance de signatures. D'eux-mêmes, deux types de réponses existent, la réponse passive et la réponse active.

Nous allons tout d'abord étudier les modes de détection d'un IDS, avant de présenter les réponses possibles à une attaque [7].

10.1 Modes de détection :

Nous notons deux modes de détection qui sont :

- La détection d'anomalies.
- La reconnaissance de signature.

Il faut noter que la reconnaissance de signature est le mode de fonctionnement le plus implémenté par les IDS du marché. Cependant, les nouveaux produits tendent à combiner les deux méthodes pour affiner la détection d'intrusion.

10.1.1 La détection d'anomalies :

Cette approche est connue aussi par l'approche de détection d'anomalies. Le but cette technique est la prédiction de comportement. Pour cela, elle utilise une base de données des comportements normaux des utilisateurs, d'un groupe d'utilisateurs, des services ou d'un système entier pour constituer un profil. Des attaques inconnues peuvent donc être détectées. Quand un comportement s'éloigne trop du comportement normal, une alarme se déclenche. Néanmoins, cet éloignement ne signifie pas forcément un comportement hostile, ce qui semble générer un taux élevé de fausses alarmes. La création du profil peut se faire grâce à plusieurs métriques: Le principe de cette approche est de considérer tout comportement n'appartenant pas au modèle de comportement normal comme une anomalie symptomatique d'une intrusion ou d'une tentative d'intrusion.

L'approche comportementale possède un certain nombre d'avantages et d'inconvénients :

- **Les avantages :**

Détection d'intrusions inconnues possibles.

L'analyse comportementale n'exige pas des connaissances préalables sur les attaques.

La détection de la mauvaise utilisation des privilèges.

- **Les inconvénients :**

Les approches comportementales produisent un taux élevé des alarmes de type faux positif en raison des comportements imprévisibles d'utilisateurs et des réseaux.

Un utilisateur peut changer lentement de comportement dans le but d'habituer le système à un comportement intrusif.

Ces approches nécessitent des phases d'apprentissage pour caractériser les profils de comportement normaux.

L'obligation d'un temps d'apprentissage pour réaliser le profil des utilisateurs .Man des utilisées, la charge CPU, les heures de connexions…etc.

10.1.2 La reconnaissance de signatures :

Ce type d'IDS contient une base de données des signatures d'attaques et essaye de faire correspondre une donnée, obtenue par les sources d'informations du système, avec celles connues. Ces signatures sont les caractéristiques d'une attaque, c'est à dire l'empreinte d'une attaque connue, et peuvent varier d'un système à l'autre.

Il est aisé de comprendre que ce type d'IDS est purement réactif; il ne peut détecter que les attaques dont il possède la signature. De ce fait, il est nécessaire de faire des mises à jour quotidiennes, de plus, ce système est aussi bon que l'est la base de signature. Si les signatures sont erronées ou incorrectement conçues, l'ensemble du système est inefficace.

Cette approche possède d'avantages et d'inconvénients qui sont comme suit [10]:

- **Les avantages :**

Prise en compte des comportements exacts des attaquants potentiels.

L'analyse basée connaissance est très efficace pour la détection d'attaque avec un taux très bas d'alarmes de type faux positif.

Les alarmes générées sont significatives.

- **Les inconvénients :**

Ne permet pas de détecter des attaques inconnues.

La base de données de la signature besoin d'une mise à jour régulière pour détecter de nouvelles attaques.

Le risque que l'attaquant à influence la détection après la reconnaissance des signatures.

Une fois une attaque détectée, un IDS a le choix entre plusieurs types de réponses, que nous allons maintenant détailler.

10.2 Réponses actives et passives :

10.2.1 Réponse passive :

La réponse passive d'un IDS consiste à enregistrer les intrusions détectées dans un fichier de log qui sera analysé par le responsable de sécurité. Certains IDS permettent de logger l'ensemble d'une connexion identifiée comme malveillante. Ceci permet de remédier aux failles de sécurité pour empêcher les attaques enregistrées de se reproduire, mais elle n'empêche pas directement une attaque de se produire.

10.2.2 Réponse active :

La réponse active, au contraire a pour but de stopper une attaque au moment de sa détection. Pour cela nous disposons d'une reconfiguration du firewall. La reconfiguration du firewall permet de bloquer le trafic malveillant au niveau du firewall, en fermant le port utilisé ou en interdisant l'adresse de l'attaquant. Cette fonctionnalité dépend du modèle de firewall utilisé, tous les modèles ne permettent pas la reconfiguration par un IDS. Dans le cas d'une réponse active, il faut être sûr que le trafic détecté comme malveillant l'est réellement, sous peine de déconnecter des utilisateurs normaux.

En général, les IDS ne réagissent pas activement à toutes les alertes. Ils ne répondent aux alertes que quand celles-ci sont positivement certifiées comme étant des attaques. L'analyse des fichiers d'alertes générés est donc une obligation pour analyser l'ensemble des attaques détectées.

11 Détecter un IDS :

Comme nous l'avons déjà signalé, il est très dangereux que la présence d'un IDS soit remarquée par un pirate. Car dans ce cas, il tentera d'obtenir un maximum d'informations sur l'IDS installé (ex. : la version utilisée) pour pouvoir l'outre passer et attaquer sans se faire remarquer.

Voici quelques techniques qui permettent de détecter un IDS :

Usurpation d'adresse MAC : les NIDS mettent l'interface de capture en mode promiscuité (promiscues mode), il est donc possible de détecter l'IDS en envoyant par exemple un ICMP « echo request » à la machine soupçonnée d'être un NIDS avec une adresses MAC inexistante.

Si la machine répond alors elle est en mode promiscuous et peut donc être un NIDS.

Mesure des temps de latence : puisque l'interface est en mode promiscuous, les temps de réponse sont plus longs. Voici une méthode pour exploiter ces temps de latence :

- le pirate génère une série de pings vers l'adresse à tester, puis il mesure et note les temps de réponse.
- le pirate sature ensuite le réseau en broadcast dans le but de ralentir l'IDS, qui recevra tous les paquets. Enfin, le pirate réémet la même série de pings en mesurant les nouveaux temps de réponse. S'ils sont bien plus élevés que les premiers temps obtenus, il est fort possible que la machine soit en mode promiscous.

Exploiter les mécanismes de réponses actives : les IPS réagissent à certaines attaques (fermer session, bloquer port…), mais en faisant cela, ils laissent souvent des empreintes (header des paquets) permettant d'identifier le type d'IPS.

12 Mesures de performance d'un IDS :

Il y a un ensemble des facteurs pour améliorer la performance d'un IDS pour traiter une grande quantité de données et minimiser le nombre de fausses alarmes [11].

La matrice de confusion présentée dans le tableau suivant est utilisée pour visualiser, pour chaque classe, les vraies classifications et les classifications prédites.

		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 1.1: Matrice de confusion

Les vrais négatifs ainsi que les vrais positifs correspondent à un fonctionnement correct du système, ce qui signifie que l'IDS a prédit avec succès respectivement le comportement normal et les attaques. Les faux négatifs sont des attaques incorrectement prédites comme des comportements normaux.

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.

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

Taux de détection :

Mesure le taux des attaques détectées par un IDS dans uenvironnement donné et pendant une durée donnée. C'est le nombre des attaques détecté sur le nombre des attaques existants dans le corpus.

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

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.

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

Conclusion :

Dans ce chapitre, nous avons présenté un aperçu sur la sécurité informatique dans un réseau et l'importance de la mise en place d'une politique de sécurité en traçant les besoins et les objectifs voulus afin de remédier aux menaces constantes que subi un réseau informatique.

Les systèmes de détection d'intrusion sont des outils permettant aux administrateurs de systèmes d'information d'être alertés en cas d'intrusion ou même de tentative d'intrusion afin de pouvoir réagir de manière adéquate.

La détection d'intrusion se fait en deux modes, détection en temps réel et différé, celle en temps réel est la plus importante puisque elle permet de réagir dans l'immédiat contre les intrusions avant que les fichiers ou les données soit copie suivie d'une destruction.

Chapitre 2

Internet des objets (IoT)

Chapitre 02 : Internet des Objet IoT

Introduction :

Aujourd'hui, Internet se transforme progressivement en un Hyper Réseau, comme un réseau formé par des multitudes de connexions entre des Artefacts (physiques, documentaires), des acteurs (biologiques, algorithmiques), des écritures et des concepts (linked data, metadata, ontologies), appelé «Internet of Things (IoT) Internet des objets (IdO)», connectant des milliards d'êtres humains, mais aussi des milliards d'objets. Il devient l'outil le plus puissant jamais inventé par l'homme pour créer, modifier, et partager les informations. Cette transformation montre l'évolution du réseau d'internet : d'un réseau des calculateurs vers à un réseau d'ordinateurs personnels, et puis vers un réseau nomade intégrant les technologies des communications. Les développements des technologies Machine-to Machine (M2M) pour le contrôle de machine à distance et aussi l'apparition dans l'année 2000 d'IP (Internet Protocole) sur les réseaux mobiles cellulaires ont accéléré l'évolution de M2M vers l'IOT.

1 Définition d'internet des objets :

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 [14].

2 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 [15]. 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, 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 [15].

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 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 [15].

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.

3 La Motivation :

Actuellement, le développement des applications IoT a été intégré dans de nombreuses tâches dans notre vie quotidienne.

Exemple :

La santé peut coûter cher, que les soins soient prodigués par un système public ou privé. La réduction des coûts est donc un mouvement commun à l'ensemble du secteur.

Dans ce cadre, les objets connectés peuvent servir à réduire certains éléments de dépenses pour les remplacer par d'autres .L'apport de l'Internet des objets peut ainsi permettre de :

- Favoriser l'hospitalisation à domicile.
- Réduire les erreurs médicales.

Optimiser la consommation de médicaments ou encore leur prise régulière (via des piluliers connectés) et encourager la prévention de certaines maladies. Le champ d’utilisation est large et pourraient permettre de générer des gains de temps [16].

4 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 prochaines années 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 [15]:

4.1 Les *Villes* Intelligentes :

Beaucoup de grandes villes ont été soutenues par des projets intelligents, comme Séoul, New York, Tokyo, Shanghai, Singapour, Amsterdam et Dubaï. Les villes intelligentes (voir Figure 2.1) peuvent encore être considérées comme des villes de l'avenir et la vie intelligente, et par le taux d'innovation de la création de villes intelligentes d'aujourd'hui, il sera devenu très faisable pour entrer la technologie IoT dans le développement des villes.

La demande exige une planification minutieuse à chaque étape, avec l'appui de l'accord des gouvernements, citoyens à mettre en œuvre la technologie d'Internet des objets dans tous les aspects. Par l'IoT, les villes peuvent être améliorées à plusieurs niveaux, en améliorant les infrastructures, en améliorant les transports.

Figure 2.1 : représente les constituants d'une ville intelligente

4.2 Le Smart Grid :

L'un des domaines d'application de l'IoT est le secteur de la distribution d'énergie intelligente, dit « Smart Grid » (voir figure 2 .2). En France, ERDF est très actif dans le développement de ce domaine, où un besoin clair en récupération d'information à différents points du réseau électrique est devenue nécessaire pour une 15 meilleure intégration des différentes sources d'énergies et une meilleure gestion de la distribution jusqu'aux utilisateurs finaux. [26]

Figure 2.2 : représente les constituants d'une smart grid

4.3 Les Appareils Intelligents :

Des appareils intelligents (voir figure 2.3) dans les soins de santé sont utilisés pour stocker et gérer les paramètres de soins clés et pour gérer les données sur les maladies capturées. Ils sont principalement déployés pour fournir des solutions de conditionnement physique en suivant les activités ciblées et des dispositifs de diagnostic utilisés pour stocker des données de dispositifs. Principalement, ils sont utilisés comme des solutions de fitness pour suivi des activités du patient et des appareils de diagnostic intelligents tels que les dispositifs de tension matérielle, les podomètres, Google verre, etc. utilisé pour capturer les données des capteurs, pour une analyse plus approfondie par le médecin.

Figure 2 .3 : représente des appareilles intelligent

4.4 Le Système De Santé Electronique :

L'internet des objets a rapidement transformé la prestation de soins. Les équipements et les capteurs sont de plus en plus « intelligents » et génèrent toujours plus de données nécessaires aux équipements médicaux, aux professionnels et profitant ainsi aux patients, en réduisant les coûts et en améliorant leur satisfaction. Les données ainsi collectées facilitent, adaptent, améliorent, anticipent ou réorganisent les soins des patients. Dans le contexte de généralisation du traitement médical électronique, l'Internet des objets est fondamental. En effet, la conception d'un système intelligent de prise de décision clinique, matérialisé par le stockage des données collectées sur les patients et leur accessibilité universelle, procurerait au médecin un excellent appui durant la phase de traitement (voir figure 2.4). L'internet des objets trouve donc tout son intérêt dans le domaine médical, et qui aussi peut améliorer le développement dans ce dernier.

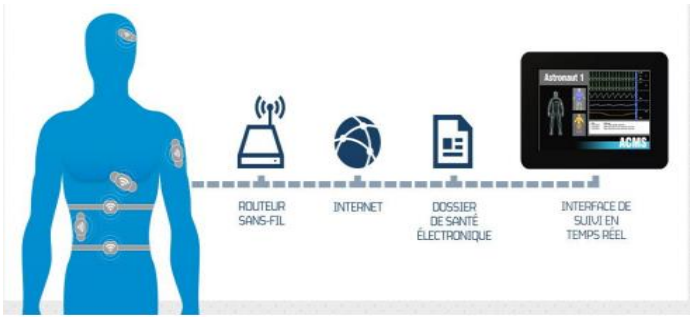

Figure 2.4: représente un système de santé électronique

4.5 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.

4.6 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.

4.7 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.

5 Infrastructure de communication:

Dans leur majorité, les objets ne se connectent pas directement à internet mais via une passerelle ou un hub numérique. L'infrastructure réseau la plus souvent mise en œuvre peut se représenter comme montrer la figure 2.5 :

Les objets connectés communiquent entre eux, ou avec les serveurs de traitement, prioritairement par le réseau Internet. Néanmoins, les types de liaison peuvent différer suivant les situations et les environnements : des objets peuvent partager directement entre eux des informations sur leur environnement pour interagir sans échanger avec des serveurs ; ou bien peuvent avoir besoin d'agréger les données avant de les transmettre à un serveur de traitement. Ils peuvent aussi être dans des zones à faible couverture avec un débit faible ou très éloignés d'un point d'accès réseau et dans ce cas utilisé un réseau dédié [17].

Dans l'infrastructure de communication, les éléments à prendre en compte concernent :

Le type de communication (courte ou longue portée).

La couverture du réseau.

La consommation énergétique de l'objet.

Le volume de données transmises.

La fréquence de captation.

La fréquence de transmission.

Le prix des capteurs/émetteurs.

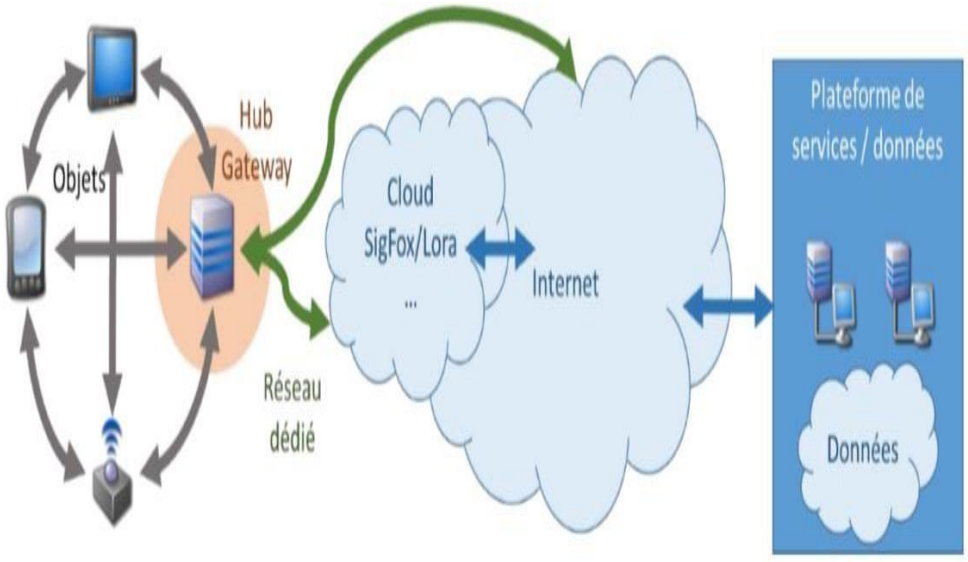

Figure2.5: Infrastructure réseau communément mise en œuvre pour les objets Connectés

6 L'évolution d'IOT et son impact dans le monde:

En 2003, la population mondiale s'élevait à environ 6,3 milliards d'individus et 500 millions d'appareils étaient connectés à Internet. Le résultat de la division du nombre d'appareils par la population mondiale (0,08) montre qu'il y avait moins d'un appareil connecté par personne. Selon la définition de Cisco IBSG, l'IoT n'existait pas encore en 2003 car le nombre d'objets connectés était relativement faible, En raison de l'explosion des Smartphones et des tablettes, le nombre d'appareils connectés à Internet a atteint 12,5 milliards en 2010, alors que la population mondiale était de 6,8 milliards. C'est ainsi que le nombre d'appareils connectés par personne est devenu supérieur à 1 (1,84 pour être exact) pour la première fois de l'histoire.

En ce qui concerne l'avenir, Cisco IBSG estime que 25 milliards d'appareils seront connectés à Internet d'ici à 2015 et 50 milliards, d'ici à 2020. Il est important de noter que ces estimations ne tiennent pas compte des progrès rapides d'Internet ni des avancées technologiques, mais reposent uniquement sur les faits avérés à l'heure actuelle. En outre, le nombre d'appareils connectés par personne peut sembler faible, mais il ne faut pas oublier que le calcul porte sur l'ensemble de la population mondiale, dont une grande partie n'est pas encore connectée à Internet. Si l'on se base uniquement sur la population disposant d'une connexion à Internet, le nombre d'appareils connectés augmente considérablement. Par exemple, nous savons qu'environ 2 milliards de personnes utilisent actuellement Internet. Si l'on utilise ce chiffre, le nombre d'appareils connectés par personne passe à 6,25 en 2010, au lieu des 1,84 précédemment indiqués [18].

La figure suivante montre l'augmentation des objets connectes par rapport le monde humain.

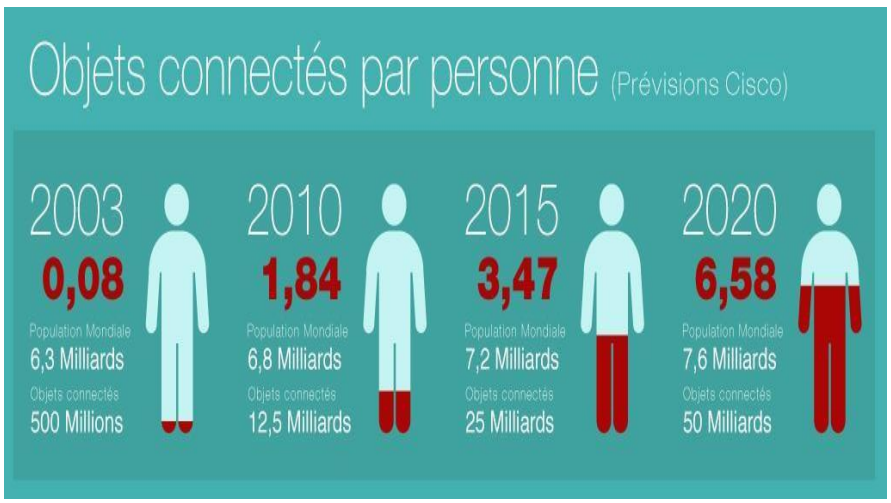

Figure 2.6: la prévision « cisco» sur les objets connectés par personne [29]

7 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 fabriquant.

Organiser : rend l'IoT plus accessible et moins conceptuel. Nous présentons maintenant les différentes architectures qui ont été proposées par certains chercheurs :

7.1 Architectures à trois et cinq couches :

L'architecture la plus élémentaire est une architecture à trois couches, 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 spécifiques 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 [19].

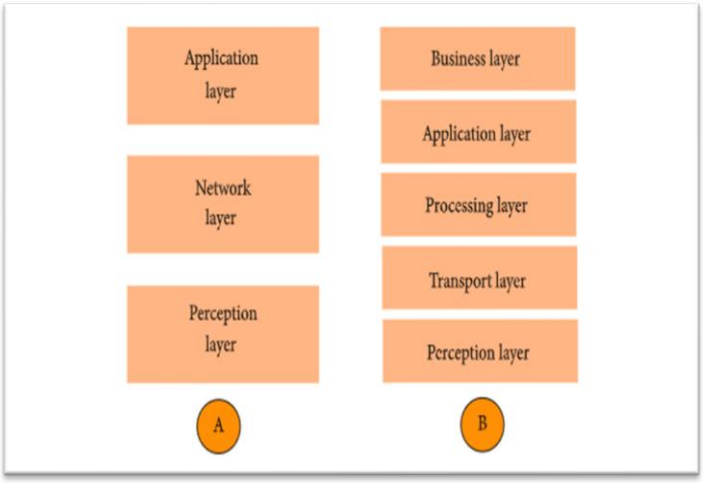

Figure 2.7 : Architecture de l'IoT (A : trois couches) (B : cinq couches) [20]

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 méga données.

La couche métier gère l'ensemble du système IoT, y compris les applications, les modèles commerciaux et de profit et la confidentialité des utilisateurs. La couche métier sort du cadre de cet article. Par conséquent, nous n'en discutons pas plus avant [19].

7.2 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 2.8, 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 [19] .

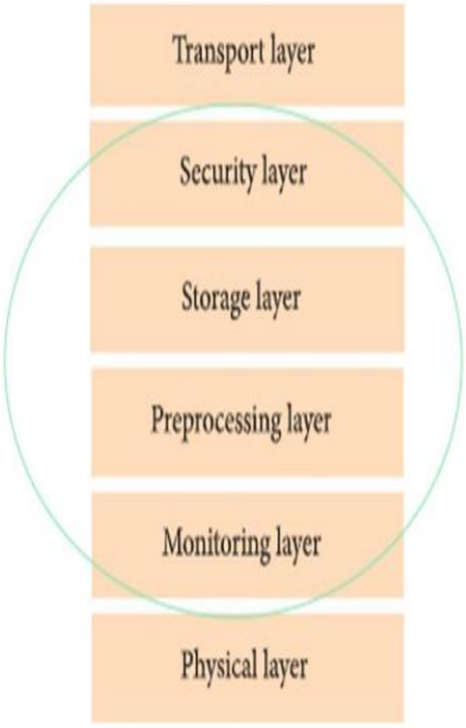

Figure 2.8 : Architecture de brouillard d'une passerelle IoT intelligente **[21]**

8 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.

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 [21]:

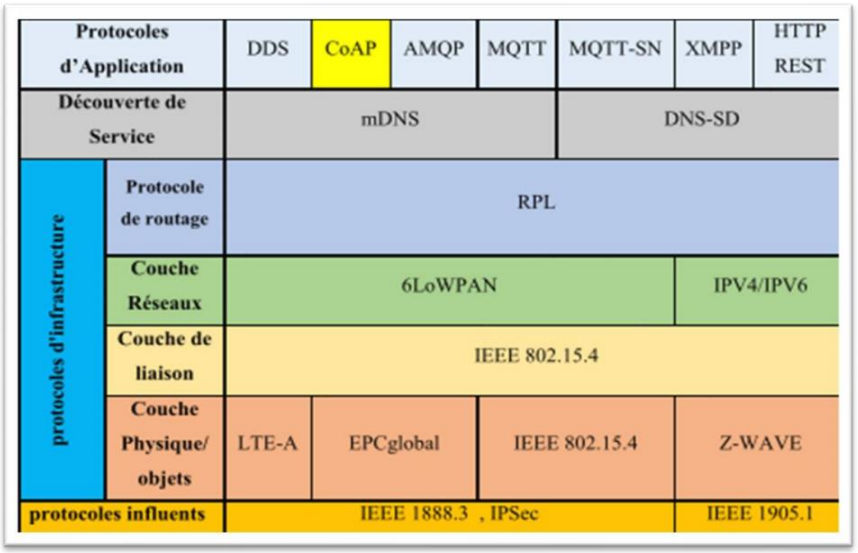

Figure 2.9: Protocoles de communication de l'internet Des Objets

9 Protocoles d'infrastructure :

9.1 6LoWPAN :

Pour pouvoir parler de protocole 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 réseauIPv6[22].

9.2 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) (Association, 2011).

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 [15].

9.3 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

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.

9.4 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 [24] .

10 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 [25]. Ces attaques sont présentées dans le tableau 2

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 deuxParties 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 2.1 : Type d'attaques dans l'IoT

11 RPL (Routing Protocol for Low power and lossy networks-LLNS):

Le protocole RPL [26] est un protocole qui a été conçu afin de prendre en charge les exigences spécifiques de ces réseaux à ressources limitées.

RPL est protocole de routage proactif à vecteur de distance qui construit un DODAG (Destination Oriented Directe Acyclic Graph) pour l'acheminement des données vers la station de base. Le DODAG construit permet à chaque nœud du DODAG de transmettre les données qu'il a récolté jusqu'au DODAG root (racine). Chaque nœud dans le DODAG sélectionne un parent selon une métrique de routage donnée et une fonction objective. Les données récoltées sont acheminéesde fils à parent jusqu'à la racine.

Les métriques de routage utilisables par RPL sont définies dans le RFC6551 . Ces métriques sont les informations qui seront prises en compte pour la création de la topologie. On calcule doncle rang d'un nœud en fonction de la métriques donnée qui peut être une mesure de la qualité d'unlien donné, de la propriété d'un nœud, et également une contrainte à respecter. On peut par exempleessayer de minimiser le délai de bout-en-bout, le nombre de transmissions nécessaires pour atteindre le puits de données, l'énergie consommée par le réseau, ou éviter d'utiliser des cheminsavec une trop forte latence, etc.

Les métriques de routage sont utilisées pour calculer le rand des nœuds à l'aide d'une fonction objectif (OF). A l'heure actuelle, deux OF sont définies : Objective Function Zero (OF0défini dans RFC6552) et Minimum Rank With Hysteresis Objective Function (MRHOF défini dans RFC6719). Le rôle de la fonction d'objectif est donc de prendre en entrée une valeur de métrique et de calculer le rang correspondant pour un nœud par rapport à un nœud donné. Elle définit aussi le processus de sélection d'un parent une fois le calcul des rangs du voisinage est effectué.

Par ailleurs, pour le bon fonctionnement du protocole RPL, chaque nœud contient la base des informations suivantes :

Un ID du nœud : cet ID (identifiant) est unique, on peut utiliser l'adresse IP du nœud parexemple.

Un rang (R) : le rang est une valeur calculée à l'aide d'une fonction OCP (Objective Code Point) où le rang d'un nœud est calculé en fonction du rand de son parent :

$$\boldsymbol{R(i) = R(p(i)) \; avec \; X(i, p(i)) + 1}$$

où le R(i) est le rang du nœud i, le p(i) est le père par défaut de i , le X(i , j) est la valeur métrique de routage utilisée. Elle pourra être la qualité de lien (ETX) ou l'énergie consommée. ETX est calculée à l'aide des rapports périodiques envoyés depuis la couche MAC vers la couche réseau sur le nombre des messages envoyés par rapport au nombre d'ACKs reçus).

Liste des prédécesseurs (ID nœud père, rang du nœud père, métrique de routage).

Père par défaut .

Liste des destinataires (ID nœud père, ID saut précèdent, métrique de routage).

11.1 Messages du protocole RPL :

Le protocole RPL comporte quatre types de messages de contrôle utilisés dans la phase de découverte de routes. Ces messages sont [27]:

DIO : DODAG Information Object (DODAGID, IDRoot, Rang du root, infos sur OCF) : envoyé de manière périodique depuis le nœud racine vers tous ses nœuds voisins.

DAO : Destination Advertisment Object (ID, Rang, IDs route infos), Le DAO est un message envoyé par les nœuds esclaves (capteurs) au nœud racine afin de répondre au message DIO.

DIS : DODAG Information Sollicitation, c'est un message de sollicitation envoyé par les nœuds non voisins du nœud racine qui n'ont reçu aucun message DIO vers l'un des voisins du nœud racine. Le nœud voisin se charge du transfert des messages DIO et DAO du et vers le nœud racine.

DAO-ACK : c'est un acquittement du message DIO envoyé par les récepteurs.

11.2 Construction du DODAG :

Le processus de construction de la structure DODAG commence à la racine ou LBR (LoWPAN Border Router) qui est généralement le nœud de collecte de données (le puits ou un actionneur). Il pourrait y avoir des racines multiples configurées dans le réseau. Le protocole de routage RPL spécifie un ensemble de nouveaux messages de contrôle ICMPv6 pour échanger des informations liées à la construction de la structure DODAG[28].

La racine commence la diffusion des informations concernant la structure en utilisant le message DIO (DODAG Information Object). Les nœuds à portée de communication de la racine recevront et traiteront ce message DIO, puis ils rendront une décision (joindre la structure ou pas) fondée sur certaines règles (selon la fonction objectif, les caractéristiques du DAG et le coût du chemin annoncé). Une fois que le nœud s'est joint à la structure, il a une route vers la racine de la structure DODAG comme la figure 2.10(a)).

La racine du DODAG est appelée le parent du nœud. Le nœud calcule son rang dans le graphe, qui représente la position du nœud dans la structure DODAG. Si ce nœud est configuré pour agir comme un routeur dans le réseau il commence à diffuser à son tour dans son voisinage les nouvelles informations de la structure qu'il vient de rejoindre à travers des messages DIOs. Sile nœud n'est pas configuré pour être un routeur alors il rejoint tout simplement la structure DODAG et n'envoie pas de message DIO. Les nœuds voisins recevant cette annonce vont répéterce processus de sélection de parent, d'ajout d'itinéraire et d'annonce des nouvelles informations concernant la structure DODAG à l'aide des messages DIOs comme montre les figures 2.10(b) ,(c). Ce processus continue jusqu'à couvrir tous les nœuds du réseau. Chaque nœud de la structure DODAG a une entrée de routage vers son parent (ou plusieurs parents selon la fonction objectif) à travers lequel ce nœud peut atteindre la racine de la structure DODAG. En outre, chaquenœud dans le graphe a un rang qui représente la position relative

de ce nœud par rapport à la racine de la structure DODAG comme montre la figure 2.10 (d). La notion de rang est utilisée par RPL à des fins diverses, y compris l'évitement les boucles.

Les différentes étapes du processus de construction graphique sont représentées dans la figure 2.10 où les messages DAO (Destination Avertissement Object) visent à maintenir les routes descendantes et ne sont utilisés que pour des applications nécessitant des trafics de type point à multipoint et point à point [29].

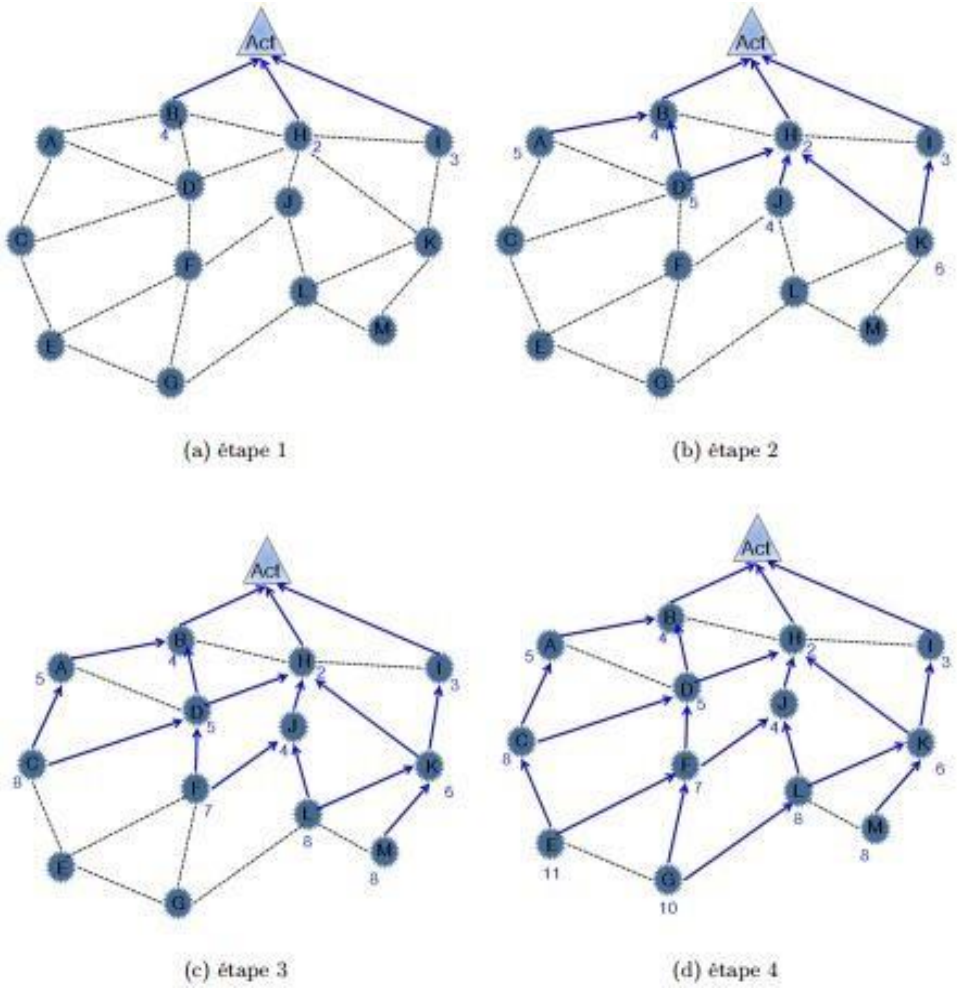

Figure 2.10: Exemple illustrant la construction d'un DODAG [9]

11.3 Trafics supportés par le DODAG :

Après la construction de la structure logique en DODAG, quand un nœud a des données à envoyer vers la racine, il les envoie vers un de ses parents (appelé son parent préféré). Ces données vont remonter la structure jusqu'à atteindre la destination finale. Ce modèle représente le modèle de trafic MP2P (multipoints à point). Les messages circulent des nœuds feuilles au(x) nœud(s) racine(s). D'autres applications nécessitent la présence d'un trafic dans le sens opposé. Ce trafic, P2MP (point à multipoints), écoule les informations vers les nœuds feuilles. Ce trafic peut provenir de l'extérieur du réseau, à partir de(s) nœud(s) racine(s). Tout cela nécessite une table de routage qui

doit être construite au niveau de chaque nœud et un mécanisme pour remplir ces routes. Ceci est accompli par le message DAO (Destination Avertissement Object). Les messages DAO sont utilisés pour annoncer l'accessibilité vers les nœuds qui peuvent être des destinations potentielles.Un nœud appartenant à la structure DODAG enverra un message DAO à son ensemble de parents. A la réception de ce message DAO, un nœud parent ajoute une entrée dans la table de routage et il envoie à son tour un message DAO à son ensemble de parents (des agrégations des informations reçues peuvent être envisagées). Ce processus se poursuit jusqu'à ce que l'information atteigne la racine du DODAG.

Le protocole RPL soutient également un autre mode appelé « mode de fonctionnement sans stockage» où aucun nœud intermédiaire ne stocke les routes vers les nœuds qui viennent de s'annoncer avec des DAOs. Le nœud puits utilise alors un routage par la source. Le protocole RPL prend également en charge le trafic de type point à point (P2P) (trafic entre deux nœudsappartenant au même DODAG). Quand un nœud envoie un message vers un autre nœud dans la structure DODAG, ce message voyage en direction de la racine du DODAG jusqu'à atteindre un nœud ancêtre commun, ayant une connaissance de la route, au niveau duquel le message sera transmis en direction de la destination finale [29].

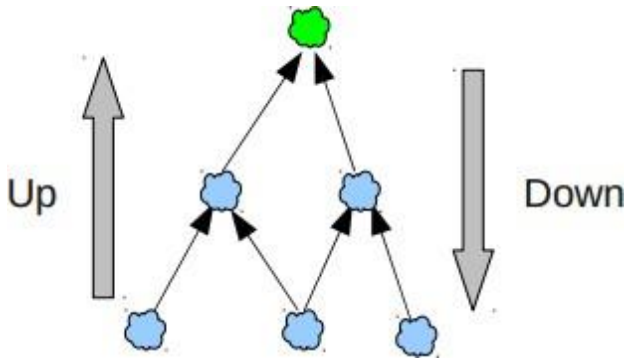

Figure 2.11: Trafics supportés par le DODAG

12 Attaques de RPL :

On propose 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 2.12 et prend en compte trois catégories d'attaquesde sécurité:

12.1 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.

12.2 La deuxième catégorie :

comprend les attaques visant la topologie du réseau RPL, 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.

12.3 La troisième catégorie :

Correspond aux attaques contre le trafic du réseau, telles que les attaques d'écoute ou le attaques de détournement.

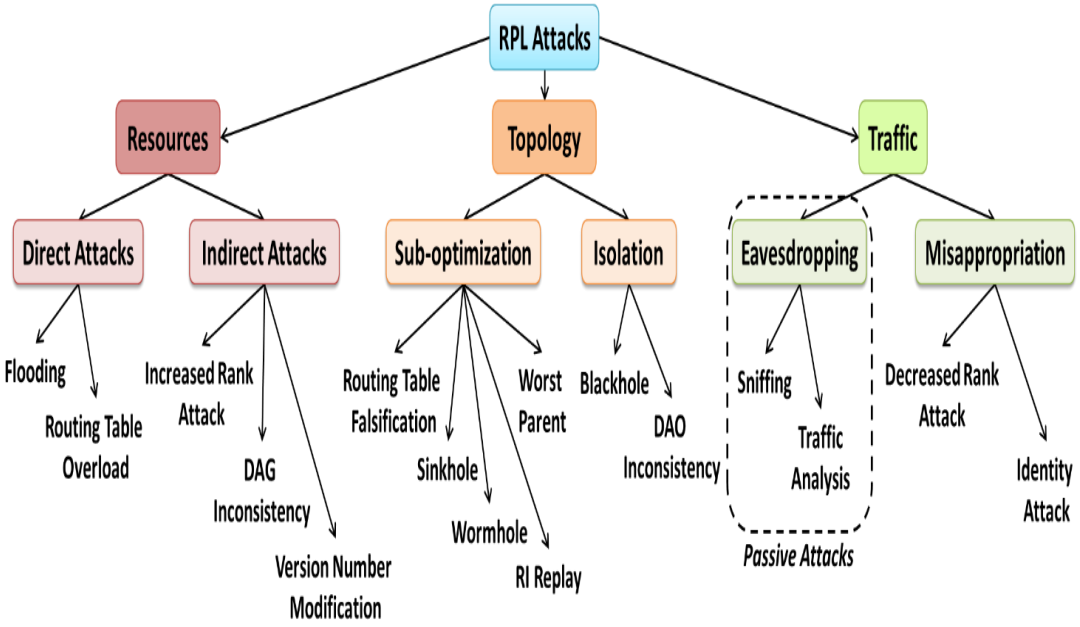

Figure 2.12 :Taxonomie des attaques contre les réseaux RPL.

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.

Dans ce chapitre nous avons présenté le protocole RPL en cours de standardisation au sein du groupe de travail de L'IETF.Nous avons représenté une description détaillée de ce protocole : son fonctionnement, la construction et le trafic supporté par le DODAG.

Chapitre 3

Techniques de classification et Calcul de Skyline

Chapitre 3 : Techniques de classification et Calcul de Skyline

Introduction :

Ce chapitre est divisé en deux partie, dans la première on va parler sur l'apprentissage automatique et quelques techniques de classifications, puis nous allons donner quelques notions sur le calcul probabiliste, présenter les réseaux bayésiens avec ses variantes ainsi que les différentes notions relatives à ces concepts.

La deuxième partie traite les requêtes Skyline qui ont reçu une grande attention dans la communauté base de données au cours des dernières décennies. Le calcul Skyline est devenu crucial pour de nombreuses applications décisionnelles multicritères.

1 Apprentissage Automatique :

L'apprentissage automatique (Machine Learning) est un domaine d'intelligence artificielle (IA) qui étudie comment les algorithmes arrivent à apprendre en étudiant des exemples. Il utilise souvent des techniques mathématiques et statistiques pour donner aux ordinateurs la capacité "d'Apprendre" (c.-à-d. : améliorer progressivement les performances sur une tâche spécifique) à partir de données.

Dans l'apprentissage automatique, au lieu d'apprendre à un ordinateur une liste massive de règles pour résoudre un problème ou réaliser une tâche, nous lui donnons un modèle avec lequel il peut évaluer des exemples, ainsi qu'un ensemble d'instructions pour mettre à jour les paramètres du modèle en vue d'améliorer la qualité de l'évaluation des exemples (i.e., réduire les erreurs d'apprentissage).

Cette technique permet l'obtention de modèles bien adaptés qui sera capable de résoudre divers problèmes (Reconnaissance de formes, traduction de texte, prédiction, ..) de manière extrêmement précise. L'apprentissage automatique est intéressant parce qu'il permet d'aborder des tâches trop difficiles à résoudre avec des algorithmes fixes écrits et conçus par des êtres humains.

Les techniques de la machine Learning sont utilisées dans de nombreux domaines, complètement distincts. À titre d'exemple, il peut s'agir du domaine médical où les machines aident à diagnostiquer les tumeurs, du domaine bancaire pour estimer la capacité d'une personne à rembourser un prêt, ou dans l'industrie du transport pour le développement de systèmes de navigation sans conducteurs.

Les algorithmes d'apprentissage automatique peuvent être catégorisés de façon générale comme : supervisés, non-supervisés, ou par renforcement, dépendent de quel genre d'expérience ils ont pendant le processus d'apprentissage. Nous introduisons [30] :

1.1 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 [31] .

1.2 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 à permettre à cet algorithme de réagir à ces informations sans supervision [32].

2 La classification bayésienne :

Modèles de représentation des connaissances, fondés sur une description graphique des variables aléatoires : Directed Acyclic Graph [33].

L'intérêt particulier des réseaux bayésiens est de tenir compte simultanément de connaissances a priori d'experts (dans le graphe) et de l'expérience contenue dans les données.

Les réseaux bayésiens sont surtout utilisés pour le diagnostic (médical etindustriel), l'analyse de risques, la détection des spam et le datamining.

Valdes et Skinner ont proposé une nouvelle approche pour la détection d'intrusions basée sur les réseaux bayésiens .Les réseaux bayésiens sont des outils de raisonnement avec des informations incertaines dans le cadre de la théorie des probabilités. Ils utilisent des graphes acycliques orientés pour la représentation des relations causales et des probabilités conditionnelles (de chaque nœud dans le contexte de ses parents) pour exprimer l'incertitude sur ces relations. Valdes et Skinner utilisent une forme simplifiée des réseaux bayésiens, appelée réseaux bayésiens naïfs, composée de deux niveaux : un nœud racine qui représente la nature de la session (normal et les différents types d'attaques), et plusieurs nœuds enfants, chacun d'entre eux correspondant à un attribut la décrivant. Les réseaux bayésiens naïfs ont plusieurs avantages dus, en particulier, à leur construction qui est très simple. L'inférence est assurée de façon linéaire En plus, la construction des réseaux bayésiens naïfs est incrémentale, dans le sens qu'elle peut facilement être mise à jour (notamment, il est toujours possible de prendre en considération de nouvelles classes) [32].

2.1 Calcul Probabiliste :

La base du calcul probabiliste est les variables aléatoires (discrètes ou continues). Une variable aléatoire est une variable qui peut prendre un ensemble de valeurs (son domaine) selon des probabilités prédéfinies (la distribution des probabilités conditionnelles). Dans ce qui va suivre nous allons considérer :

P (A) : La probabilité d'observer un événement A.

P (A, B) : La probabilité jointe d'observer deux événements A et B ensemble.

P (A|B) : La probabilité conditionnelle d'observer l'événement A, sachant la valeur de l'événement B.

Deux variables sont indépendantes si et seulement si

$$P(A|B) = P(A).\ P(B).$$

Ceci exprime le fait que l'observation de A ne dépend pas de B : quel que soit la valeur de B, nous avons toujours les mêmes probabilités d'observer A. L'indépendance est une propriété symétrique : Si A est indépendant de B, alors, automatiquement, B est indépendant de A.

La base de tout calcul probabiliste est le fameux théorème de Bayes quiexprime une probabilité jointe comme le produit d'une probabilité conditionnelle et une autre probabilité.

$$P(A, B) = p(A/B).p(B) = p(B/A).p(A)$$

Dans le cas où les deux variables aléatoires sont indépendantes le théorème de Bayes devient :

$$P(A, B) = P(A).P(B)$$

La règle de chaîne de Bayes permet d'appliquer le Théorème de Bayesrécursivement dans le cas où l'on a plus que 2 Variables. Par exemple :

$$P(A, B, C) = P(C/A, B).p(B/A).P(A)$$

Un réseau bayésien est définie par :

Un graphe oriente sans circuit G= (V, E), où V est l'ensemble des nœuds deG, et E l'ensemble des arcs de G.

Un espace probabilisé fini (Ω, Z, p).

$$P(V_1, V_2, V_3, \dots, V_n) = \sum_{i=0}^{n} P\ (V_i, C(V_i)\ )$$

Un ensemble de variables aléatoires associées aux nœuds du graphe et définies sur tel que :

Où est l'ensemble des causes (parents) de dans le graphe G.

Un réseau bayésien est donc un graphe causal auquel on a associé une représentation probabiliste sous-jacente. Cette représentation permet de rendre quantitatifs les raisonnements sur les causalités que l'on peut faire à l'intérieur du graphe [34].

L'utilisation essentielle des réseaux bayésiens est de calculer des probabilités conditionnelles

d'événements reliés les uns aux autres par des relations de cause à effet. Cette utilisation s'appelle inférence.

Une difficulté essentielle des réseaux bayésiens se situe précisément dans l'opération de transposition du graphe causal à une représentation probabiliste.

2.2 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 etles 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.

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, …, an relatives 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:

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

2.2.1 La probabilité nulle :

Le calcul des probabilités conditionnelles et a priori basé sur les fréquences, ce qui peut s'avérer entaché d'erreur si la valeur d'un attribut n'apparaît pas avec toutes les classes dans l'ensemble d'apprentissage. En effet, ceci peut entraîner des probabilités conditionnelles nulles qui vont réduire à zéro les probabilités de certaines classes.

2.2.2 Fonctionnement d'un réseau bayésienne naïve :

Inputs :

- les appels systèmes
- L'ensemble d'apprentissage.

Outputs :

- déterminer la classe du processus (normal / anormal)

Algorithme :

Apprentissage

Déterminer les fréquences d'apparition de chaque classe à partir de L'ensemble d'apprentissage. Ces fréquences sont calculées avec la formule suivante:

$$p(ci) = \frac{f(ci)}{N}$$

Sachant que : f (ci) est le nombre de fréquence de la classe. N est le nombre total du processus dans les données d'apprentissage.

Déterminer les fréquences d'apparition de chaque terme sachant la classe en utilisant la formule 1.

Sachant que : f (A/ci) est le nombre d'appel système appartenant à la classe normal ou anormal.

Inférence

Déterminer la classe du processus I= (selon la formule: 3)

Afin de classer n'importe quel nouveau processus caractérisé par ses valeurs d'attributs (appels systèmes):

Les classes choisies seront celles dont la probabilité est la plus grande [35].

3 SVM (Support vector machine) :

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 lenombre 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 [36].

4 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 d'attributs considérés commeles meilleurs parmi tous les autres au sens d'un critère donné.

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. [37]

5 Radom Forest :

Forêt aléatoire crée plusieurs arbres de décision et les fusionne pour obtenirune 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 [37].

6 Perceptrons multicouches (MLP) :

Des travaux ultérieurs avec des perceptrons multicouches ont montré qu'ils sont capables d'approximer un opérateur XOR ainsi que de nombreuses autres fonctions non linéaires.

Tout comme Rosenblatt a fondé le perceptron sur un neurone McCulloch-Pitts, conçu en 1943, les perceptrons eux-mêmes sont des blocs de construction qui ne s'avèrent utiles que dans des fonctions plus vastes telles que les perceptrons multicouches. Le perceptron multicouche est le monde de l'apprentissage en profondeur: un bon point de départ pour apprendre à utiliser l'apprentissage en profondeur. Un perceptron multicouche (MLP) est un réseau neuronal artificiel profond. Il est composé de plus d'un perceptron. Ils sont composés d'une couche d'entrée pour recevoir le signal, d'une couche de sortie qui prend une décision ou d'une prédiction concernant l'entrée, et entre ces deux, un nombre arbitraire de couches masquées qui constituent le véritable moteur de calcul du MLP. Les MLP avec une couche cachée sont capables d'approximer n'importe quelle fonction continue. Les perceptrons multicouches sont souvent appliqués aux problèmes d'apprentissage supervisé. Ils s'entraînent sur un ensemble de paires entrée-sortie et apprennent à modéliser la corrélation (ou lesdépendances)

entre ces entrées et ces sorties. La formation implique l'ajustement des paramètres, des poids et des biais du modèle afin de minimiser les erreurs [37].

7 Optimisation minimale séquentielle (SMO) :

Optimisation minimale séquentielle ou SMO. La formation d'une machine à vecteurs de support nécessite la résolution d'un très gros problème d'optimisation de la programmation quadratique (QP). SMO divise ce gros problème de QP en une série de problèmes de QP les plus petits possibles. Ces petits problèmes de QP sont résolus de manière analytique, ce qui évite d'utiliser une optimisation de QP numérique fastidieuse comme une boucle interne. La quantité de mémoire requise pour SMO est linéaire dans la taille du jeu d'apprentissage, ce qui permet à SMO de gérer des jeux d'entraînement très volumineux. Etant donné que le calcul matriciel est évité,

SMO bascule quelque part entre linéaire et quadratique dans la taille du jeu d'apprentissage pour divers problèmes de test, tandis que l'algorithme SVM de segmentation standard varie entre linéaire et cubique dans la taille du jeu d'apprentissage. Le temps de calcul de SMO étant dominé par l'évaluation SVM, SMO est donc le plus rapide pour les SVM linéaires et les fichiers fragmentés. Sur des ensembles de données clairsemés du monde réel, SMO peut être plus de 1000 fois plus rapide que l'algorithme de segmentation [37].

8 Les requêtes Skyline :

Dans un contexte décisionnel, certaines requêtes ne renvoient aucun résultat. Dans ces requêtes, l'utilisateur recherche les tuples pour lesquels les valeurs de certains critères sont optimales. C'est le caractère « multicritère » de ces interrogations qui les rend généralement infructueuses. En effet, tel tuple peut être optimal pour un critère mais pas pour un autre, il est alors éliminé du résultat alors qu'il aurait pu être pertinent pour l'utilisateur. Par exemple, si l'on considère une base de données immobilières, la recherche du logement «idéal »peut combiner des conditions sur le prix, le plus bas possible, la surface, la plus grande possible, et l'éloignement du lieu de travail, le plus réduit possible. Évidemment il est vraisemblable que ce logement idéal n'existe pas, d'où l'absence de réponse à ce typede requête. Pourtant certains logements pourraient s'avérer pertinents pour l'utilisateur parce que, situés dans une zone proche, mais non voisine, ils réunissentles critères de surface maximale et de prix minimal.

Afin d'apporter une réponse adéquate au type de requêtes décrites, l'opérateur SKYLINE [38] a été introduit. Il considère l'ensemble des critères de choix d'une recherche comme autant de préférences et extrait les tuples globalement optimaux pour cet ensemble de préférences. Ainsi plutôt que de rechercher une hypothétique solution idéale, il extrait les candidats les plus proches possibles des souhaits de l'utilisateur. Son principe général s'appuie sur la notion de dominance. Un objet ou un tuple est dit dominé par un autre si, pour tous les critères intéressant le décideur, il est moins optimal que cet autre. Un tel tuple est éliminé du résultat, non pas parce qu'il est non pertinent pour un des critères mais parce qu'il est non optimal selon la combinaison de tous les critères. En d'autres termes, il existe au moins une meilleure solution pour l'utilisateur qui, elle, sera retenue.

8.1 Définition :

Skyline est une opération importante dans de nombreuses applications de retourner un ensemble de points intéressants d'un potentiel énorme espace de données. Compte tenu d'une table, l'opération trouve toutes les lignes qui ne sont pas dominés par d'autres tuples.

8.2 Relation de dominance :

Soit C = {c1,c2,...,cd} l'ensemble des critères sur lesquels porte l'opérateur Skyline. Et Soit

deux tuples t et t′, la relation de dominance suivant l'ensemble de critères

C est définie comme suit :

$t \succ_C t' \Leftrightarrow t[c1] \leq t'[c1]$ et $t[c2] \leq t'[c2]$ et ... et $t[cd] \leq t'[cd]$

Lorsque $t \succeq_C t'$ et $\exists\, c_i \in C$ tel que $t[c_i] < t'[c_i]$, la dominance est stricte , elle est notée $t \succ_C t'$.

Lorsqu'un tuple t domine un tuple t′ (i.e. $t \succ_C t'$), cela signifie que t est équivalent ou «Meilleur » que le tuple t′ pour tous les critères choisis. Comme nous considérons que les critères sont minimisés, les valeurs de t pour tous les critères sont inférieurs ou égales à celles de t′. Ainsi dans le cadre d'une recherche multicritère les tuples dominés par d'autres (au moins un)ne sont pas pertinents et sont éliminés du résultat par l'opérateur Skyline.

8.3 L'opérateur SKYLINE :

Soit r une relation, le Skyline de r suivant C'est l'ensemble des tuples qui ne sont dominés par aucun autre, suivant l'ensemble de critères C :

$SKY_C(r) = \{t \in r \mid \nexists t' \in r, t' \succ_C t\}$

Exemple de référence : Le tableau 2 décrit un ensemble de points E correspondant à des propositions d'hôtels avec les attributs suivants : le prix de l'hôtel, la distance de l'hôtel à la plage.

Hôtel	Prix	Distance
a	**1600**	**40**
b	**2400**	**10**
c	**3000**	**50**
d	**3600**	**40**
e	**2300**	**20**
f	**3000**	**30**
g	**3600**	**40**

Tableau 3.1: L'ensemble des hôtels avec critères associés

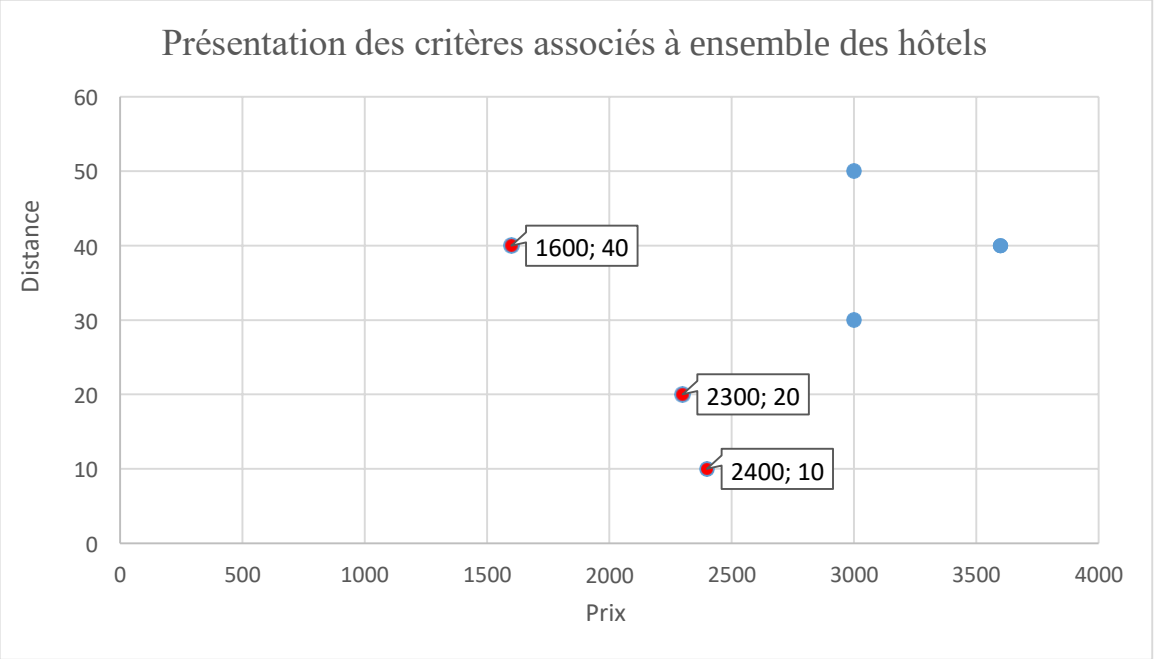

Figure 3.1: Skyline associé à ensemble des hôtels

L'ensemble des points Skyline obtenus à partir de l'ensemble des points E définis dans un espace à n dimensions D, noté Sky(D,E), est construit à partir des points qui dominent tous les autres points sur au moins une dimension (ensemble MaxSky(D,E)) mais aussi des points qui ne sont pas meilleurs sur une dimension donnée mais qui constituent une solution compromis intéressante pour l'utilisateur (ensemble CompSky(D,E)).Sky ((Prix, Distance), E)= {a, b, e} avec MaxSky ((Prix, Distance), E)= {a, b} car a et b ont les meilleures valeurs respectivement sur les dimensions 'Prix' et 'Distance et CompSky ((Prix, Distance), E)= {e} car e estmeilleur que a et b respectivement sur les deux dimensions examinées.

Conclusion

Dans ce chapitre on exposer quelque technique de classification à savoir réseaux bayésiens, SVM (support vector machine), J48, Random Forest, Perceptrons multicouches (MLP), en nous concentrant sur la classification naïve. Ensuite nous avons introduit les requêtes Skyline et expliquer le principe de calcul de Skyline.

Dans ce qui suit-on va détailler notre solution proposée et présenter nos résultats d'expérimentation.

Chapitre 4

Contribution dans la détection d'intrusion

Chapitre 4 : Contribution dans la Détection D'intrusion

Introduction :

Dans ce Chapitre, nous présentons notre modèle proposé (skyline-IoT), les différentes étapes de construction du modèle ainsi que les résultats qui en découlent, la proposition skyline-IoT se base sur deux niveaux:

Le premier niveau contient les meilleurs classificateurs, ces classificateurs ont été sélectionnés en utilisant l'opérateur Skyline en se basant les trois mesures de performances (Taux de détection et l'exactitude et taux des fausses alertes).

Le deuxième niveau est représenté par un réseau bayésien naïf pour fusionner les prédictions du premier niveau et délivrer une décision finale.

1 Description et structure de notre 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.

Premier niveau contient Cinque classificateur après application de Skyline en choisir j48 et RF Comme illustré dans la figure 4 .1.

Le deuxième niveau contient un seul classificateur utilisé parce qu'il a la possibilité d'ajouter des nœuds en tant que classificateur final, Il analyse les prédictions sélectionnées des différents meilleurs classificateurs du premier niveau plus la connexion et prend la décision finale, 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 4 .2

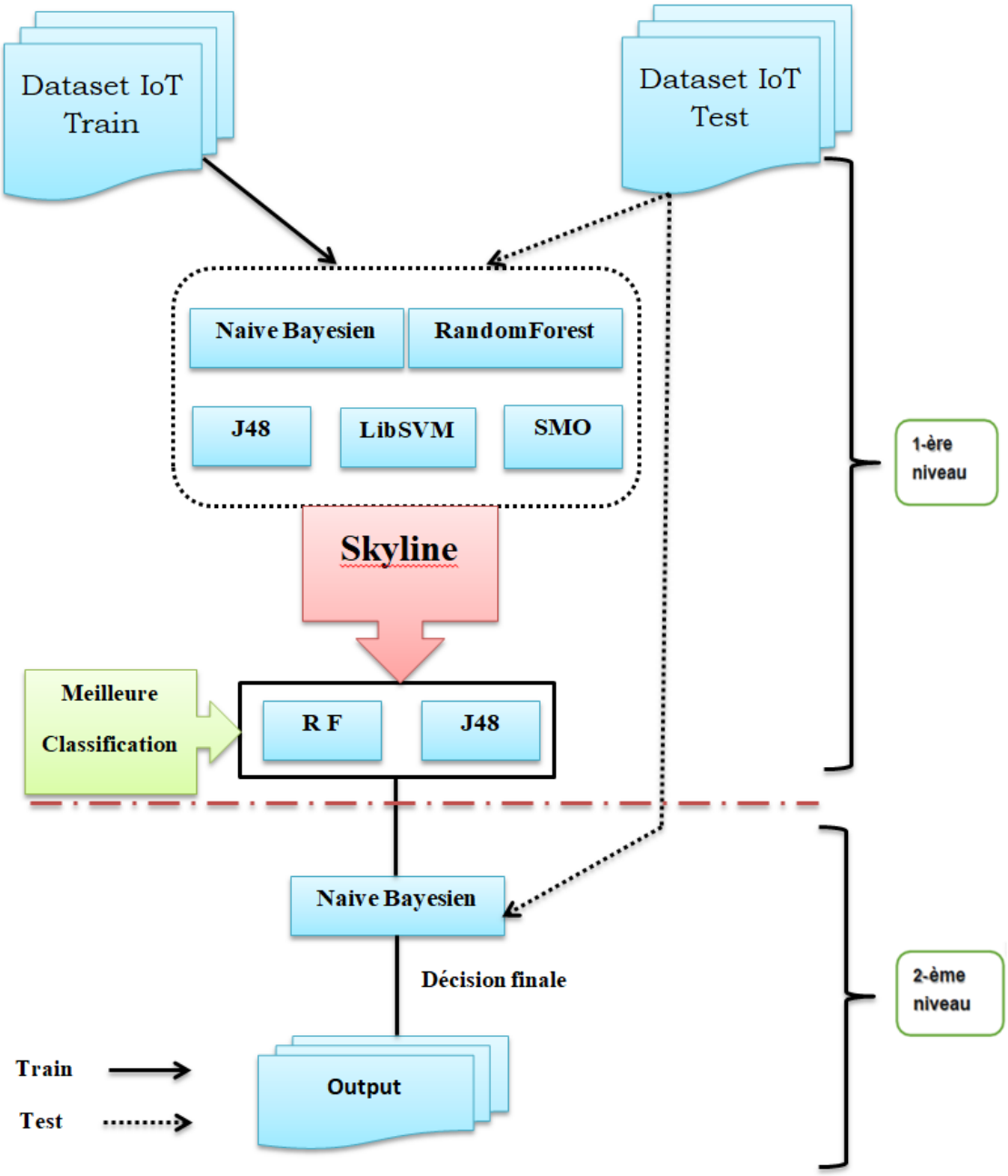

Figure 4.1 : Structure de Skyline-IoT

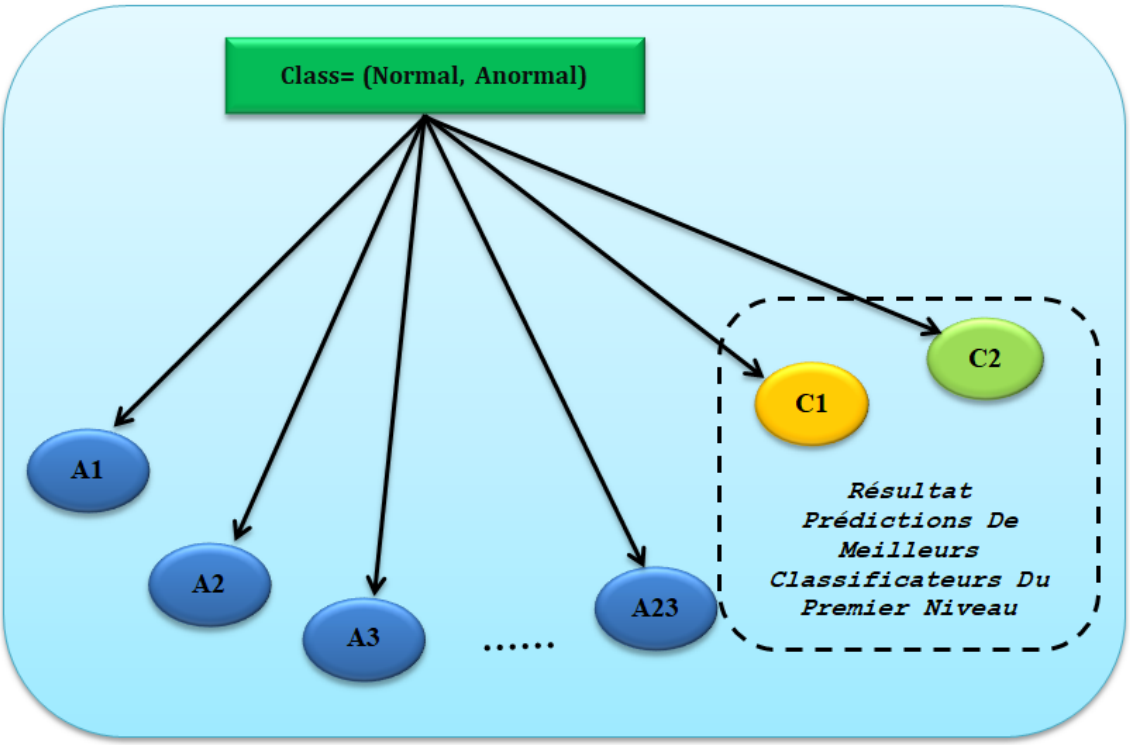

Figure 4.2 : Modèle bayésien [39]

2 Mode de fonctionnement :

La sélection de Classificateurs , la phase d'apprentissage et de test constituent les étapes qui nous ont permis de construire Skyline_IoT.

Skyline est une opération importante dans de nombreuses applications de retourner un ensemble de points intéressants d'un potentiel énorme espace de données. Compte tenu d'une table, l'opération trouve toutes les lignes qui ne sont pas dominés par d'autres tuples.

2.1 Sélection des différents classificateurs :

Dans le but de récolter les meilleurs classificateurs pour skyline_IoT, nous effectuons l'opérateur Skyline tel que les classifier qui reste ne sont dominés par aucun autre.

En effet en premier temps, on a procédé à une sélection de classificateurs les plus utilisés dans les travaux d'IDS.

Les différents classificateurs comparés sont : Naive bayesien, Random Forest (RF), SimpleVector Machine (SVM), arbre J48 (J48), Sequential minimal optimization (SMO), ces derniers présentent dans le tableau 4.1 ci-dessous

Mesures → / Classificateurs ↓	DR	Exactitude	FAR
Naivebayesien	***0,960***	***0,9620***	***0,010***
J48	***0,987***	***0,987***	***0,003***
RandomForest	***0,993***	***0,993***	***0,002***
LibSVM	***0,896***	***0,896***	***0,026***
SMO	***0,884***	***0,883***	***0,029***

Tableau 4.1: Les performances des classificateurs

Ensuite, une élimination de quelques classificateurs parmi ces derniers en appliquant l'opérateur Skyline qui consiste 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 classificateurs incomparables.

C'est ainsi qu'on retiendra les 2 classificateurs pour former le premier niveau de skyline_IoT à savoir **RF, J48** .

Mesures → / Classificateurs ↓	DR	Exactitude	FAR
Naivebayesien	***0,960***	***0,9620***	***0,010***
J48	***0,987***	***0,987***	***0,003***
RandomForest	***0,993***	***0,993***	***0,002***
LibSVM	***0,896***	***0,896***	***0,026***
SMO	***0,884***	***0,883***	***0,029***

Tableau 4.2: Skyline des classificateurs

2.2 La phase d'apprentissage :

La phase d'apprentissage nous prépare à la phase de test, et est constituée de deux niveaux :

Le premier niveau est formé avec une partie de l'ensemble de données d'apprentissage où chaque connexion représente une entrée pour le classificateur du premier niveau.

Cette partie de l'ensemble de données d'apprentissage correspondant à notre base de données d'apprentissage a été créé à partir du **Dataset Balancier**.

Le deuxième niveau est formé skyline -IoT tel que chaque classificateur donne sa prédiction par rapport à un type de connexion pour lequel ce classificateur est sélectionné, ensuite cette prédiction et le classificateur qui lui convient seront fusionnés avec l'ensemble de données de test.

2.3 La phase de test :

Nous traitons chaque enregistrement de l'ensemble de données de test par les différents meilleurs classificateurs de premier niveau, Ensuite nous utilisons les sorties des prédictions et sa classificateur comme des entrées pour le classificateur du deuxième niveau.

2.4 Description data set:

Nous avons utilisé un ensemble de données IoT [40] , Ils rééquilibraient des instances de la classe pour faciliter l'apprentissage.

Les ensembles de données d'apprentissage et de test : Nous avons créé notre ensemble de données contenant 144069 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 4.3 ci-dessous résume la distribution des attaques ainsi le comportement normal de notre ensemble de données d'apprentissage et de test.

Avant d'équilibrage de donnée		**Après l'équilibrage de donnée**
Toutes les catégories	48024	144069
Rank	9367	29818
Hello	5046	28746
Version	3196	27778
Black	1493	28805
Normal	28922	28922

Tableau 4.3 : jeu de données d'origine et jeu de données sur échantillonné

2.5 Description de différents 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 S)
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 4.4: Description de différents attributs

3 Expérimentation :

Nous avons effectué une série d'expérimentation avec le dataset qui représente l'ensemble de données de détection d'intrusion le plus utilisé dans la dernière année, Weka est utilisé pour la mise en œuvre des différents classificateurs, Les résultats sont obtenus sur un PC Dell Intel(R) Core™) i5-6300 Cpu @ 2,50 GHz, et 08Go de RAM à l'aide de weka 3,9,5.

Nous procédons ensuite à l'expérimentation du premier niveau en utilisant Naive bayesien , Random Forest, , SVM, J48, SMO suite à l'analyse de leur performances.

Dans le deuxième niveau, nous avons utilisé le Naïves Bayes pour sa caractéristique probabiliste simple et apte à intégrer les prédictions et ses classificateurs du premier niveau.

4 Etude comparative :

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

Mesures → / Classificateurs ↓	**DR**	**Exactitude**	**FAR**
J48	0,987	0,987	0,003
RandomForest	0,993	0,993	0,002
LibSVM	0,896	0,896	0,026
SMO	0,884	0,883	0,029
Skyline _IOT	0,998	0,998	0,0004

Tableau 4.5: La comparaison entre les classificateurs

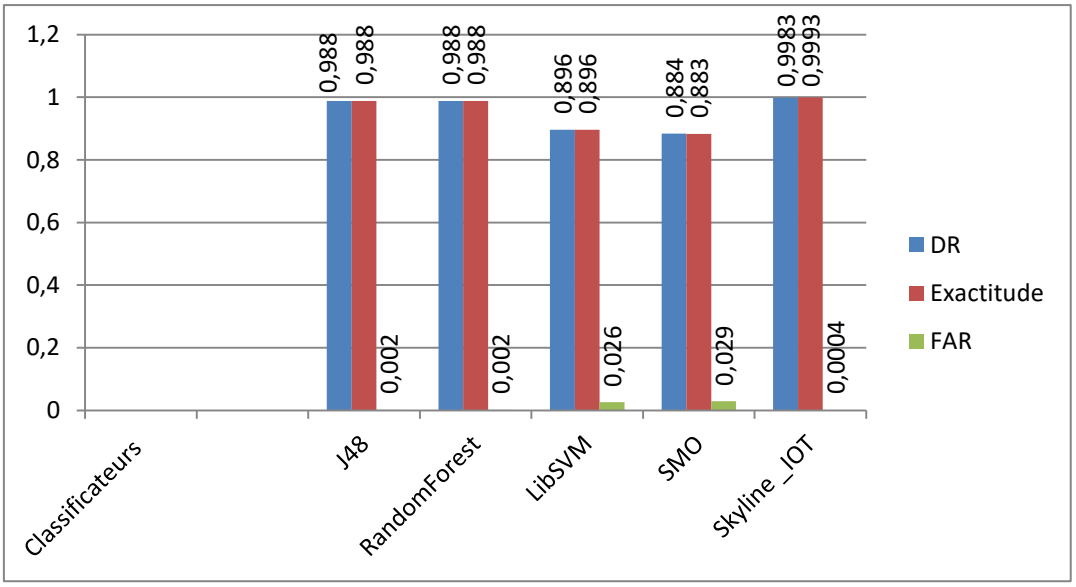

Figure 4 .6 :Etude comparative entre les classificateurs

4.1 Discussion

Pour évaluer les performances de skyline_IoT, nous avons comparé ses performances avec des travaux connexes [41] et [42] , Le résultat de cette étude comparative est résumé dans le tableau suivant .

Mesures → / Classificateurs ↓	Précision %	Recall %	F1-Score %	Accuracy %
Travaux connexe[41]	95,7	95,7	95,7	/
Travaux connexe[42]	99,4	99,3	/	99,33
Skyline _IOT	99,8	99,8	99,8	99,8

Tableau 4.6: La Comparaison entre skyline-IOT et travaux connexes

Comme le montre Tableau 4.6 , skyline_IoT a montré sa haute performance pour le taux d'exactitude(Accuracy) le plus élevé, et un fort taux de détection(Recall) , Skyline_IoT est

plus précis que les autres utilisés dans cette étude comparative avec un taux d'exactitude égal 99.8 % et précision 99,8% ,f1-score 99,8%

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.

L'idée principale dans ce travail est l'hybridation de plusieurs techniques de classification (l'utilisation de l'opérateur Skyline pour choisir les meilleurs classificateurs) et fusionner les différents prédictions en utilisant un réseau bayésien naïf.

Comme cité plus haut nous avons montré la haute performance de notre modèle par rapport aux travaux connexes, notre modèle donne un taux d'exactitude le plus élevé avec le taux de fausse alarme le plus bas et un bon taux de détection, 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 un réseau IoT.

Chapitre 5

Réalisation et implémentation

Chapitre 5 : Réalisation et implémentation

Introduction

Après avoir présenté l'architecture de skyline_IoT dans le chapitre précèdentet 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'environment de développement, Nous détaillons le processus de l'implémentation ainsi que la principale interface qui le compose à travers des fenêtres de capture.

1 Les Outils de la réalisation

1.1 Weka :

Weka est un logiciel d'exploration de données qui utilise une collection d'algorithmes d'apprentissage automatique, Ces algorithmes peuvent être appliqués directement aux données ou appelés à partir du code Java [36].

Weka est une collection d'outils pour :

- Régression
- Clustering
- Association
- Data pre-processing
- Classification
- Visualisation

1.1.1 Description globale de Weka

Weka (Waikato Environment for Knowledge Analysis) est un environnement de fouille de données développé par le groupe de recherche "machine Learning" du département d'informatique de l'université de Waikato en Nouvelle-Zélande, Il est utilisé dans le domaine de la recherche, de l'éducation et de l'industrie, Il est écrit dans le langage Java et testé sur plusieurs plateformes tels que Linux et Windows, Cet environnement est un logiciel "open source" et est disponible sur le site du groupe de recherche "machine learning" du département d'informatique de l'université de Waikato [43].

Weka est une collection d'algorithmes d'apprentissage dont le but est de réaliser des tâches de fouille de données, Les algorithmes peuvent être appliqués directement à un ensemble de données ou appelés via un programme Java, Weka contient les outils pour le prétraitement de données, la classification, la régression, le groupement (clustering), les règles d'association et la visualisation,

En effet, Weka permet d'effectuer un prétraitement sur un ensemble de données, d'appliquer un algorithme d'apprentissage, et d'analyser les résultats et les performances d'un classificateur, Il est aussi bien adapté pour intégrer de nouveaux algorithmes d'apprentissage.

La version utilisée dans cette Mémoire est la plus récente : la version 3,9,5 (figure 5.1).

Figure 5.1 Environnement Weka

1.1.2 Composants de l'environnement Weka

Weka possède plusieurs composants à savoir :

- **Explorer** : ce module regroupe tous les packages importants de Weka à savoir le prétraitement, les algorithmes d'apprentissage, le groupement (clustering), les associations, la sélection des attributs et la visualisation.
- **Experimenter** : permet d'exécuter plusieurs algorithmes d'apprentissage en mode lot (batch) et de comparer leurs résultats.
- **KnowledgeFlow environment** : fournit les mêmes fonctionnalités que le composant "Explorer", Ces fonctionnalités sont représentées sous forme graphique et sont utilisées pour construire un schéma de flux de connaissances via une interface drag-and-drop.

1.1.3 Préparation de l'ensemble des données d'apprentissage et de test (dataset)

Cette opération est requise pour l'utilisation de tout outil de l'environnement Weka et c'est pour cette raison que l'on a préféré la décrire avant d'entamer la description des outils de Weka, Généralement, d'après la littérature, les données servant à l'expérimentation dans n'importe quel domaine sont stockées dans des fichiers Excel et le format des données utilisées par les tâches de fouille de données est "arff", Avant d'appliquer donc les algorithmes d'apprentissage, ces données doivent être transformées en un fichier de données avec l'extension "arff" pour être lu soit par *Explorer*, *Experimenter* ou *KnowledgeFlow*, Les transformations consistent à :

- Sauvegarder le fichier de données d'apprentissage et de test au niveau de l'outil Excelsous la forme ,csv (comma separated value);
- Lire le fichier avec l'extension "csv" dans un éditeur;
- Ajouter en entête du fichier les informations nécessaires au module "Explorer";
- Définition des données d'apprentissage par une relation sous forme de : @relation nom-des-données-d'apprentissage Dataset (figure 5. 2),
- Définition des attributs du dataset avec leur type, e,g,, @attribute (figure 5.2),
- Fermeture de ces ajouts par le mot réservé @data,
- Transformer toutes les virgules par le point et tous les points virgule par la virgule, Cesdeux petites tâches doivent être faites dans l'ordre;
- Sauvegarder le fichier transformé des données sous l'extension "arff".

Voici un exemple de fichier de données sous le format "arff",

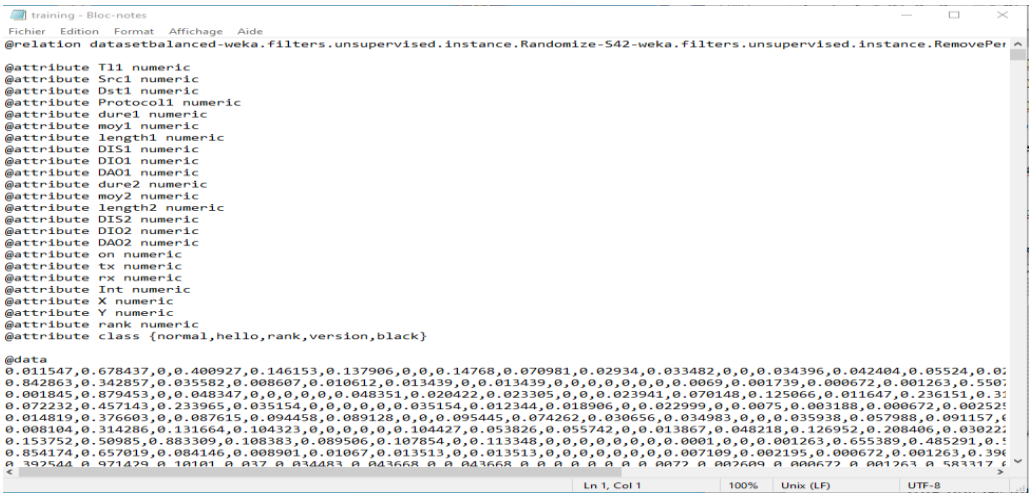

Figure 5.2 Extrait du fichier ",arff"

Une fois les transformations effectuées, le processus de fouille de données est déclenché, Il est illustré via un exemple utilisant le dataset de la collection des données de l'environnement *Weka*.

1.1.4 Explorer

Cet outil permet à un utilisateur de réaliser certaines étapes du processus de fouille de données à savoir le prétraitement, l'application d'un algorithme d'apprentissage et l'analyse, Pour ce faire, *Explorer* regroupe plusieurs packages tels que les filtres, les classificateurs, les clusters, les règles d'association, la sélection des attributs et un composant de visualisation, Cet outil permet donc de réaliser les expérimentations, en utilisant des moyens tels que les classificateurs, les clusters et les règles d'association, qui nécessitent au préalable un traitement des données, Dans ce document, on se concentre uniquement sur le prétraitement et l'application d'un algorithme d'apprentissage (classificateur), Ce prétraitement est présenté et une illustration du fonctionnement de l'outil *Explorer* est fournie,

1.1.5 Phase de prétraitement

Cet outil peut être activé en cliquant sur l'option *Explorer* de la figure 5.1, La sélection de cette option fait apparaître la fenêtre représentée par la figure 5.3.

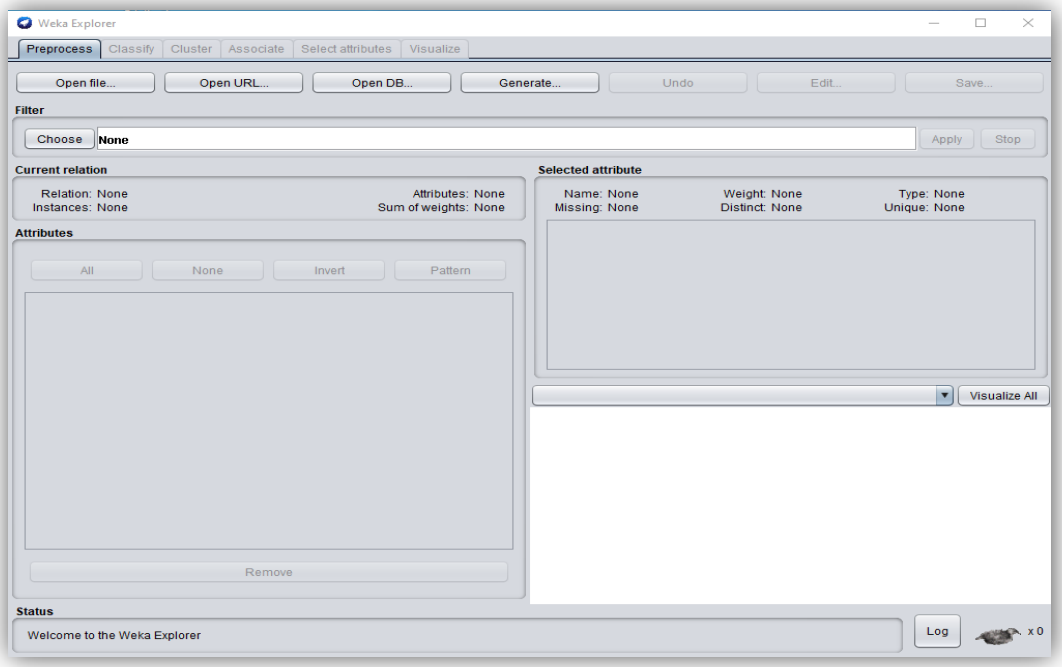

Figure 5.3 Fenêtre Explorer

Dans la fenêtre de prétraitement, nous trouvons :

- Dans la partie supérieure, les trois possibilités d'ouverture d'un fichier de données d'apprentissage qui sont "Open file", "Open URL" et "Open DB";
- Le choix du filtre en cas de nécessité, d'information sur la relation et l'attribut sélectionné, d'une partie où tous les attributs sont affichés et une zone de visualisation en forme d'histogrammes de la distribution de l'attribut sélectionné,

La figure 5.4 nous montre le contenu de la fenêtre de prétraitement après chargement du fichier de données de l'expérimentation via l'activation de l'option "Open file",

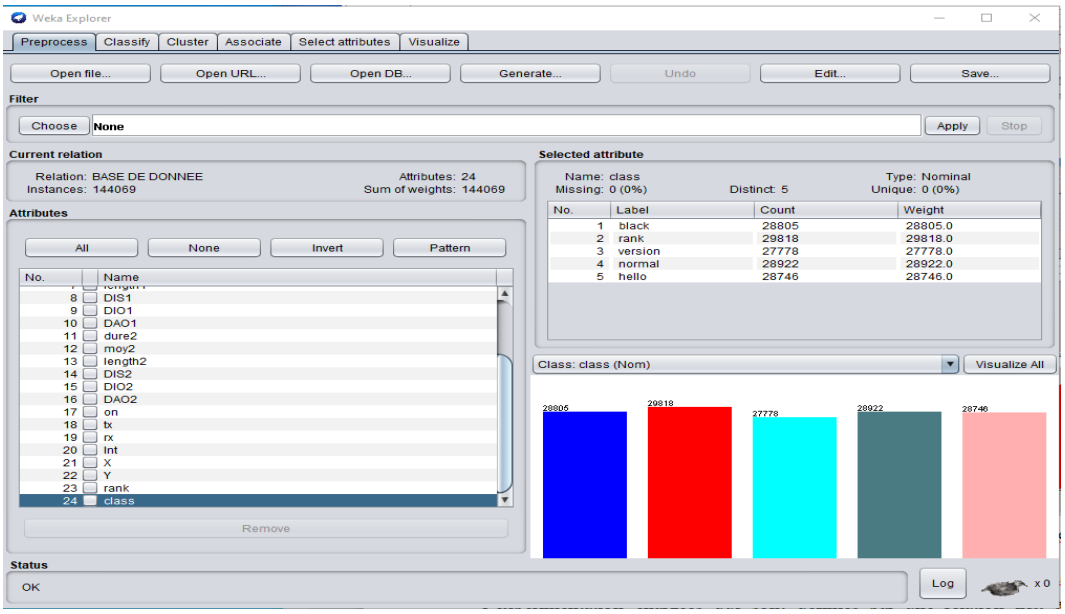

Figure 5.4 Chargement de l'ensemble des données

Dans la partie gauche, nous avons dans le premier cadre les informations sur les données de l'expérimentation chargées qui sont définies par une relation, ayant 144069 instances et 24attributs , Tous les attributs sontaffichés dans le cadre attributs, La sélection de la variable *class* dans la liste des attributs fait apparaître dans la partie droite les informations sur cette variable tels que le nom et le type, Avec la sélection de la variable à visualiser dans la partie droite, un histogramme des valeurs prises est affiché,

1.1.6 Application d'un algorithme d'apprentissage

Avant d'appliquer un algorithme d'apprentissage sur des données, un certain nombre de choix doivent être faits :

- choisir l'algorithme d'apprentissage.
- choisir le type de test :

- utiliser l'ensemble d'apprentissage comme ensemble de test.
- fournir l'ensemble de test.
- appliquer la validation croisée avec le choix du nombre de partitions (fold) de l'ensemble des données (dans nos exemples, on a opté pour ce choix).
- appliquer "percentage split" (partage de l'ensemble de données en un ensemble de données d'apprentissage et un ensemble de données test) ;
- le cas par défaut est 2/3 des données pour l'apprentissage et 1/3 des données pour le test.
- sélectionner d'autres options si nécessaire.

Sélectionner la variable à prédire (class pour notre exemple schématisé par les différentes figures).

- lancer l'exécution de l'algorithme en activant le bouton "start".

La figure suivante montre les résultats obtenus par l'application de l'algorithme J48 sur le dataset, Les résultats sont structurés en trois volets : un volet résultats sommaires, un volet résultats par classes et un volet matrice de confusion, Les résultats sommaires donnent le nombre total d'instances classifiées correctement et incorrectement, la valeur de kappa, l'erreur absolue moyenne, l'erreur racine carrée moyenne, l'erreur relative absolue et l'erreur racine carrée relative, Le choix d'un ou de plusieurs de ces critères dépend de l'intérêt que leur porte l'utilisateur final, Les résultats par classes nous fournissent le taux d'instances classifiées correctement et incorrectement via les taux TP (true positif) et FP (false positif), la précision et d'autres informations statistiques, Quant à la matrice de confusion, cette table nous donne plus de précision concernant le nombre d'instances correctement et incorrectement classifiées correspondants aux taux TP et FP pour chaque classe, Les éléments de la diagonale de la matrice représentent les instances correctement classifiées et les autres éléments des colonnes sont les instances faux positifs, Par contre, la somme des nombres pour chaque ligne donne le nombre d'instances de chaque classe .

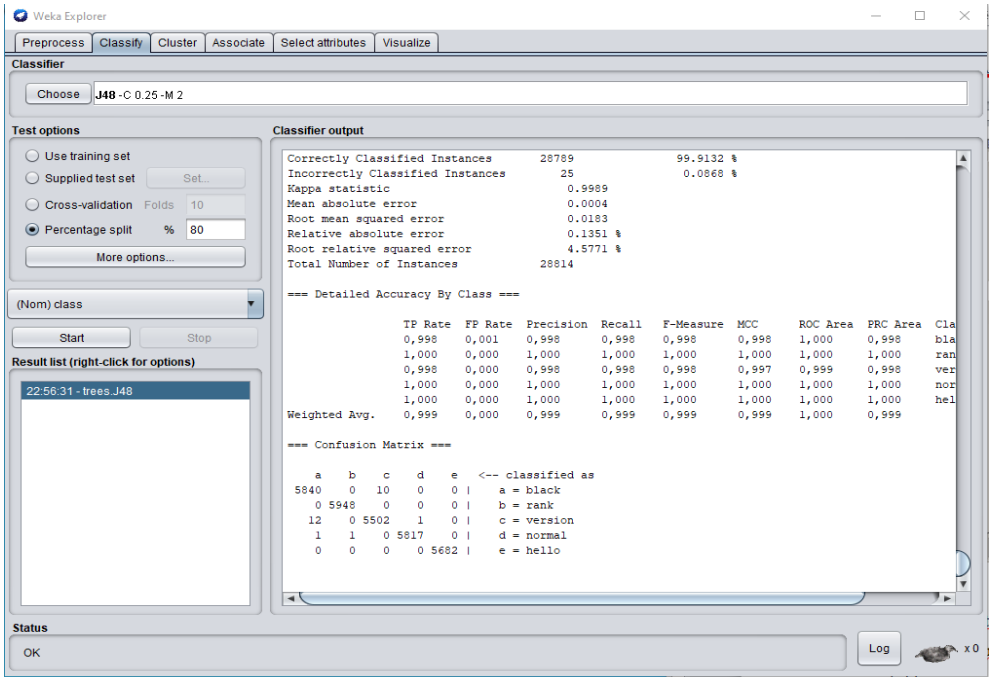

Figure 5.5 Résultats de l'algorithme J48

D'autres possibilités sont offertes en cliquant sur le bouton droit de la souris et en pointant sur le nom de l'algorithme dans le champ "result list" (partie gauche), La figure 5.6 présente toutes les possibilités.

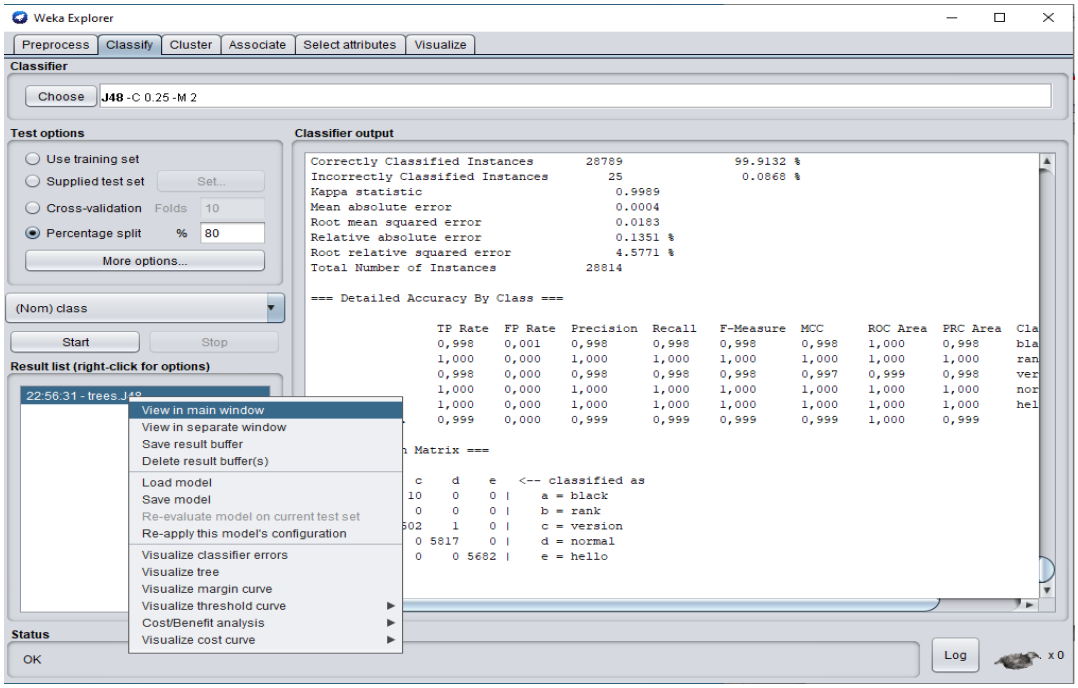

Figure 5.6 Les possibilités de visualisation

1.2 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 :

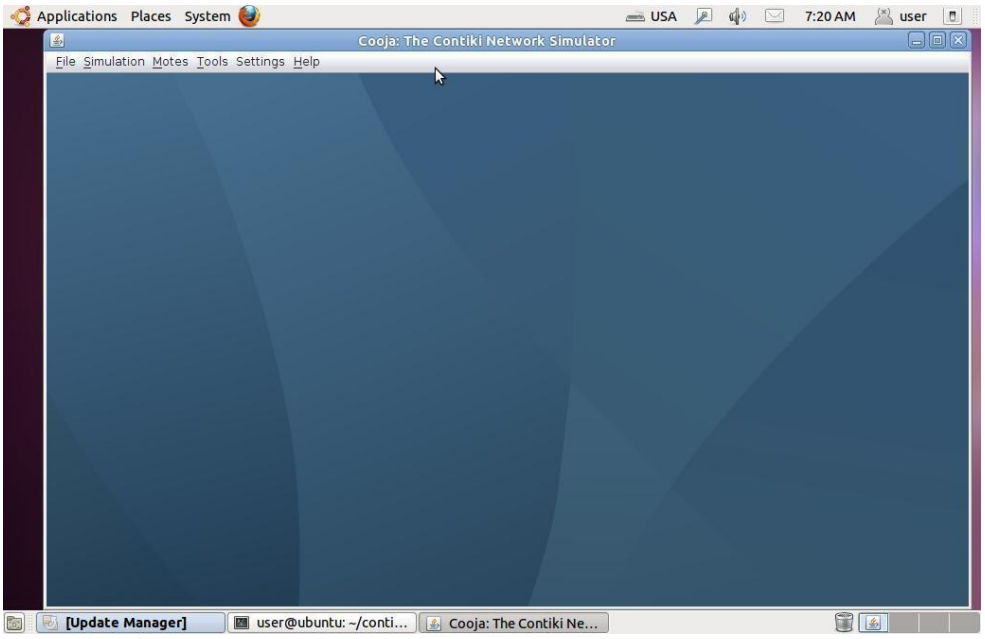

Figure 5.7 : Premier affichage Cooja

On clique sur File (Fichier), ensuite sur New Simulation (Nouvelle simulation) et l'écran illustré à la Figure 5.8 s'affiche à nouveau.

Il n'est pas nécessaire de modifier les paramètres de cet écran.

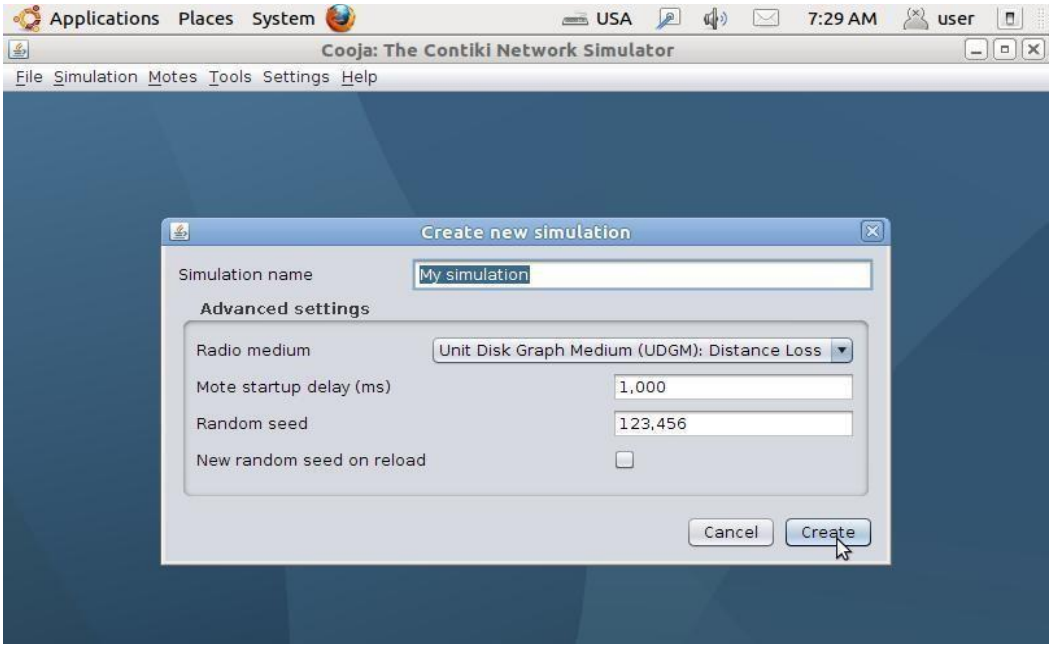

Figure 5.8 : Création d'une nouvelle simulation

Le bouton Créate permet de lancer l'écran de simulation initial, comme le montre la Figure 5.9

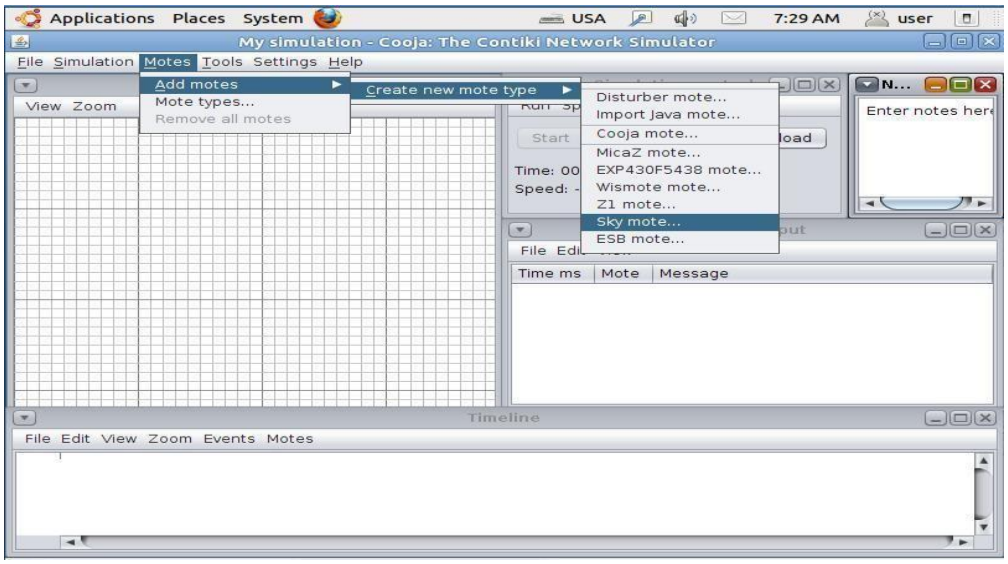

Figure 5.9 : Écran initial de simulation Cooja

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 5.10.

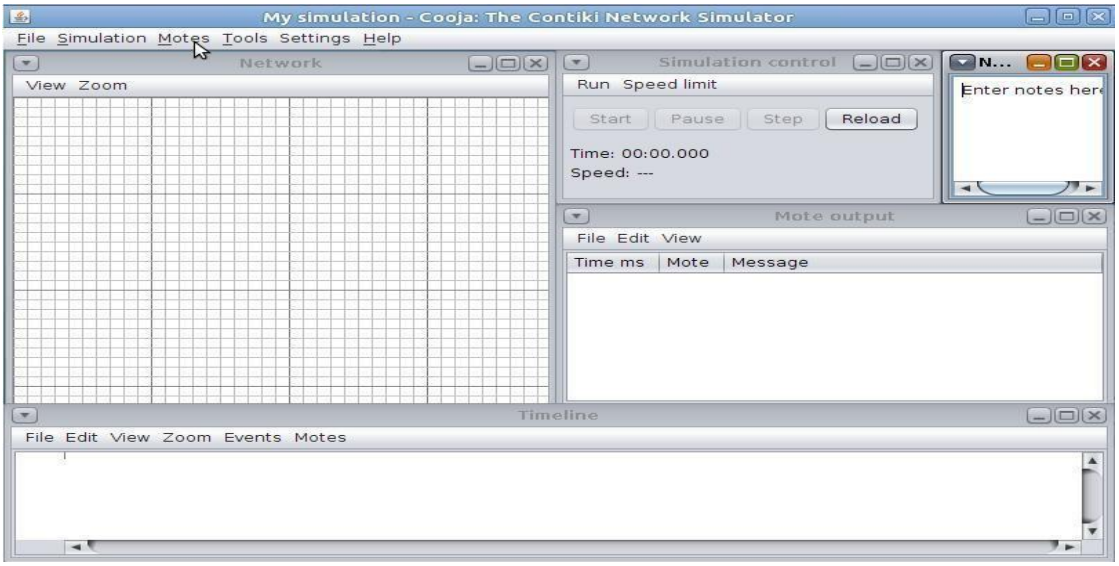

Figure 5.10 : Ajouter Motes

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 5.11.

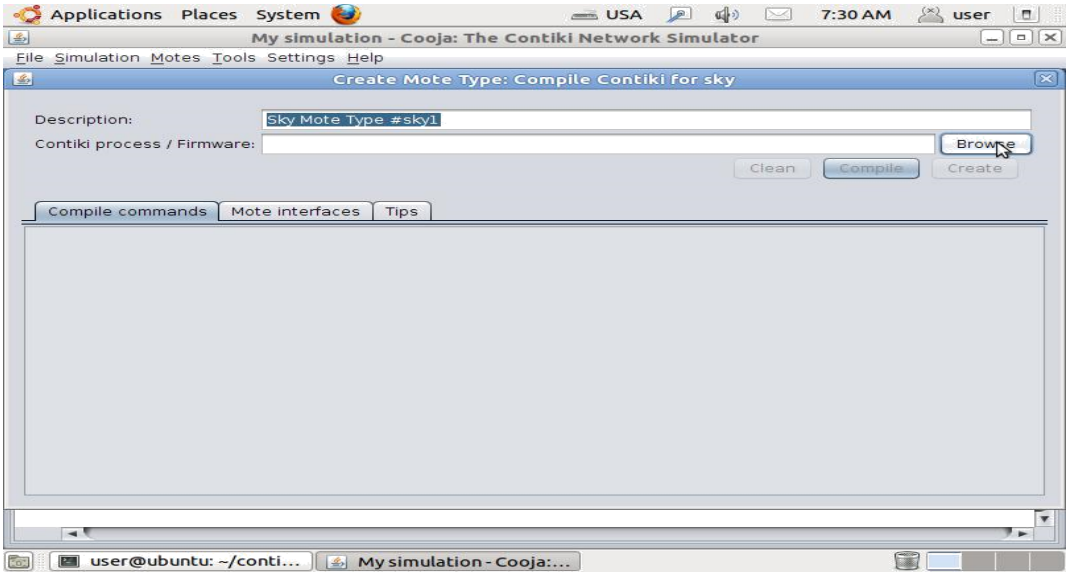

Figure 5.11 : parcourir le Mote

Comme on peut le distinguer dans la Figure 5.11, 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 lefirmware en langage C du mote qui va maintenant être créé.

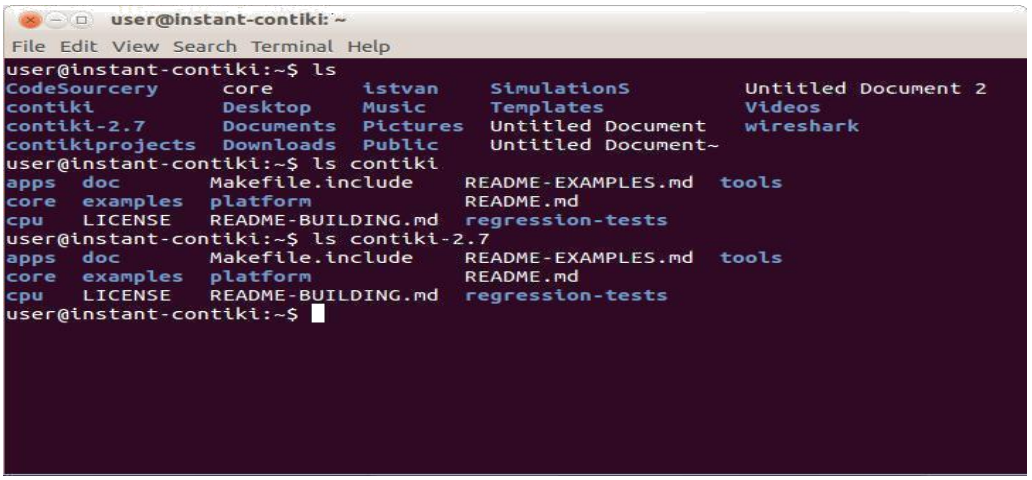

Figure 5.11 : Les fichiers contiki

On clique 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 5 .12 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.

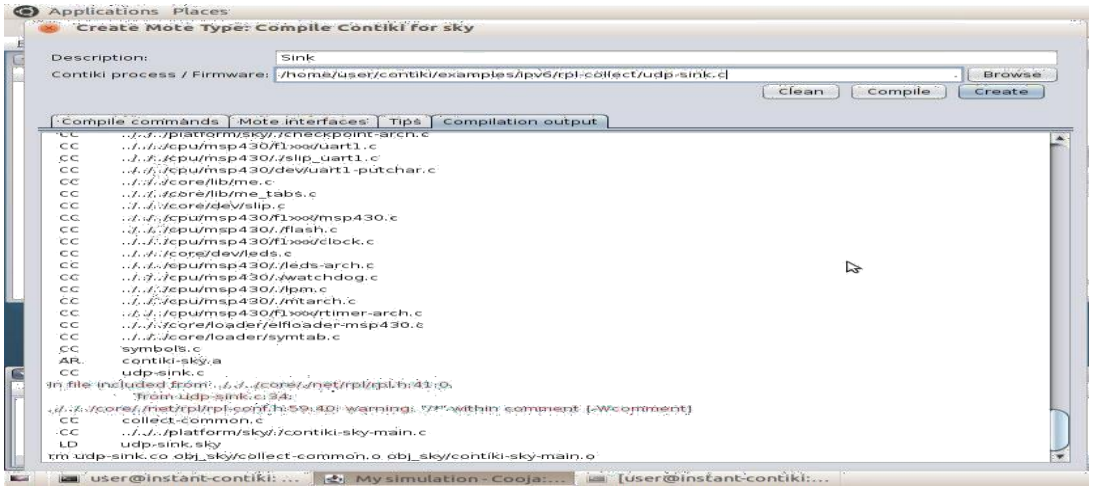

Figure 5.12 : Compilation de Mote Cooja

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 5.13

Comme il s'agit d'une Mote Sink simplement un 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 :

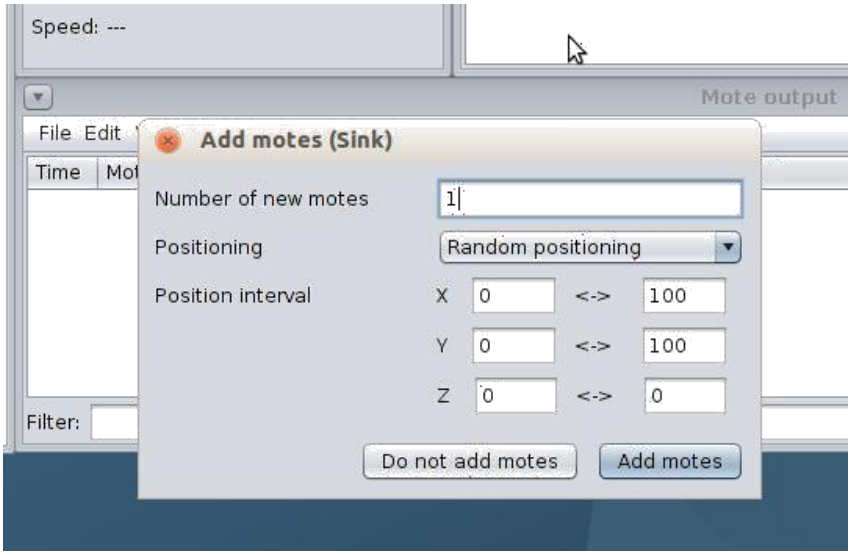

Figure 5.13 : Ajouter des Motes Cooja

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.

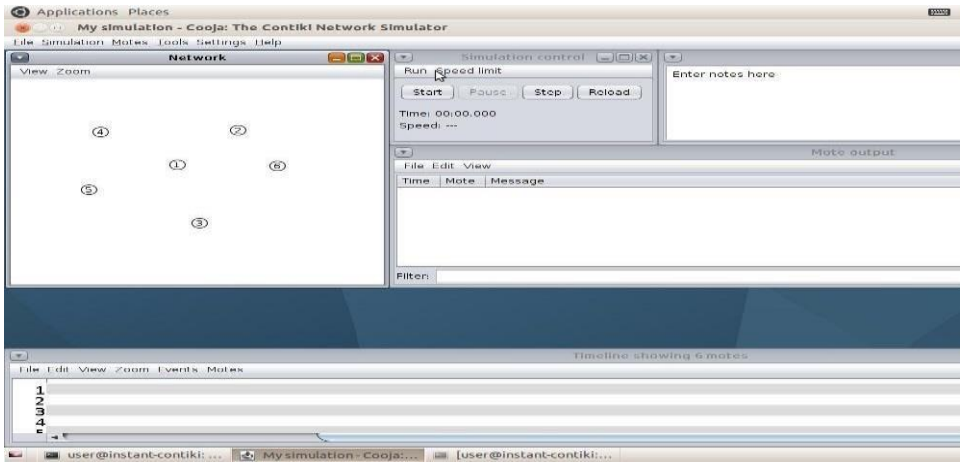

Figure 5.14 : Topologie initiale crée

Conclusion

Dans ce chapitre on a présenté quelques outils nécessaires à la réalisation (WEKA,…) afin de démontrer la capacité de notre système et en fusionner les 5 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 Générale :

Avec l'augmentation du nombre d'appareils connectés à l'Internet des objets, leur sécurité devient le premier obstacle. Lorsque nous parlons de l'Internet des objets, cela signifie des données sensibles partout, et facilement accessibles. Il existe de nombreuses recherches dans le domaine de la sécurité des réseaux Iots, mais peu d'entre elles correspondent à l'environnement réel de l'Internet des objets et aux scénarios réels auxquels ils sont exposés.

Dans ce travail, nous avons étudié l'impact des attaques de routage (RPL) dans un réseau IoTs, 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 IoTs pour obtenir un bon IDS. L'objectif de ce travail est de construire un système de détection d'intrusion contre les attaques de routage dans l'Internet des Objets basé sur des algorithmes d'apprentissage automatique. Pour former notre modèle, nous avons utilisé un jeu de données d'attaques de routage. Ce jeu de données a été construit avec le simulateur de COOJA (des informations capturées durant les communications simulées entres les nœuds d'un réseau IoTs et le nœud malveillant dans un réseau RPL), la base de données contient quatre attaques principales (blackhole ,decreased rank , version number, hello flood). Il contient également des caractéristiques importantes telles que la position géographique du nœud et l'énergie.

Dans cette optique, nous avons proposé un modèle hybride d'un IDS comportemental basée sur l'intégration des décisions d'un ensemble de classifieurs dans un réseau bayésien. Notre modèle proposé est composée de deux niveaux. Le premier niveau contient des classifieurs choisies en se basant sur l'opérateur Skyline (pour retourner les meilleurs classifieurs selon les trois mesures de performance), et le deuxième niveau est représenté par un réseau bayésien naïf pour retourner la décision finale en prenant en compte les décisions des classifieurs du premier niveau.

Ce modèle nous a donné des résultats satisfaisants quant à la détection des attaques rares mais aussi dans la classification rigoureuse des attaques des autres catégories ainsi une solution aux problèmes majeurs d'un IDS comportemental qui sont le taux élevé de faux positifs et une meilleure détection des attaques.

Au vu de ces résultats, nous avons comme perspective de tester ce modèle sur d'autres bases de données plus récentes et de choir les attributs pertinents lors des phases d'apprentissage afin d'accroitre la précision du modèle.

Bibliographie

[1] Laurent Bloch Christophe Wolfhugel, Sécurité informatique, Principes et méthode à l'usage des DSI, RSSI et administrateurs, Eyrolles, 2eme édition, Mise en œuvre d'une solution de sécurité basé sur les IDS, Juin 2014

[2] Laurent Poinsot «Introduction à la sécurité informatique», support de cours, Université Paris 13.

[3] les virus informatique clusif 2005, page 10

[4] Les virus et les spam, Page 37 (https://www.sophos.com)

[5] Philippe Biondi, Architecture expérimentale pour la détection d'intrusions dans un système informatique, Article de recherche, Avril-Sptembre 2001

[6] Laurent Bloch-Christophe Wolfhugel. Sécurité informatique .EYROLLES, 2eme edition. 2005.

[7] Le grand livre de la sécurité informatique. Securite Info, Editions du 6 nouvembre 2006.

[8] Desgeorge, G. (2000). La sécurité des réseaux. Cour.

[9] L .Me, V. Alano, Détection d'intrusion dans un système informatique : méthodes et outils, Article de recherche, Détection d'intrusion dans un système informatique : application au centre de calcul, 2013/2014.

[10] Thierry Evangelista, Les IDS Les systèmes de détection d'intrusions informatiques édition DUNOD.

[11] Thème Vers un nouveau système de détection d'intrusions basé sur l'approche comportementale, 2016-2017.

[12] DABOUR, I., & HADJI, I. (). Etude et mise en place d'un système de détection prévention d'intrusion (IDSIPS) réseau. Etude de cas SNORT.

[13] Bellovin, S. M. (1999). Distributed firewalls. Login.

[14] Benghozi, P.-J., Bureau, S., & Massit-Folea, F. (2008). L'Internet des objets. Quels enjeux pour les Européens

[15] Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of things: A survey on enabling technologies, protocols, and applications. IEEE.

[16] https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/,consulte janvier 2022

[17] Un 360° pour bien les comprendre, Décembre 2016.

[18] Dave Evans," L'Internet des objets Comment l'évolution actuelle d'Internet transformet-elle le monde ?" Cisco Internet Business Solutions Group (IBSG), Avril 2011.

[19] Sethi, P., & Saran gi , S. R. (2017). Internet of things: architectures, protocols, and applications. Journal of Electrical and Computer Engineering, 2017.

[20] Bonomi, F., Milito, R., Natarajan, P., & Zhu, J. (2014). Fog computing: A platform for internet of things and analytics. In Big data and internet of things: A roadmap for smart environments (pp. 169–186). Springer

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

[22] Ko, J., Terzis, A., Dawson-Haggerty, S., Culler, D. E., Hui, J. W., & Levis, P. (2011). Connecting low-power and lossy networks to the internet. IEEE Communications Magazine, 49(4), 96– 101.

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

[24] Vasseur, J., Agarwal, N., Hui, J., Shelby, Z., Bertrand, P., & Chauvenet, C. (2011). RPL: The IP routing protocol designed for low power and lossy networks. Internet Protocol for Smart Objects (IPSO) Alliance, 36.

[25] Limbasiya, T., & Doshi, N. (2017). An analytical study of biometric based remote user authentication schemes using smart cards. Computers & Electrical Engineering, 59, 305–321 .

[26] S. Dawson-Haggerty, A. Tavakoli and D. Culler, "Hydro: A Hybrid Routing Protocol for LowPower and Lossy Networks," 2010 First IEEE International Conference on Smart Grid Communications, Gaithersburg, MD, 2010, pp. 268-273.

[27] JP. Vasseur, M. Kim, K. Pister, N. Dejean, D. Barthel, Routing Metrics Used for Path Calculation in Low-Power and Lossy Networks, IETF Request for Comments 6551, March 2012.

[28] P Levis, T.Clausen, J.Hui, O.Gnawali,J.Ko, "The Trickle Algorithm", IETF Request for Comments 6206, March 2011.

[29] B.Romdhani, "Exploitation de l'hétérogénéité des réseaux de capteurs et d'actionneurs dans la conception des protocoles d'auto-organisation et de routage", Thèse de doctorat, en informatique, Institut National des Sciences Appliquées de Lyon , juillet 2012.

[30] Une Approche de Deep Learning pour la Recommandation des Services Web, 2018, pp 25.

[31] Une Approche de Deep Learning pour la Recommandation des Services Web, 2018, pp 26.

[32] https://www.lemagit.fr/definition/Apprentissage-non-supervise.

[33] https://perso.liris.cnrs.fr/alain.mille/enseignements/Master_PRO/TIA/RBayesiens /Intro_RB.pdf.

[34] Naim P., P.H.Wuillemin, P.Leray, O.Pourret, A.Becker. Réseaux bayésiens, Eyrolles, Paris, 2007, Système de détection d'intrusion basé sur la classification comportementale des processus ,2011/2012.

[35] http://dspace.univmsila.dz:8080/xmlui/bitstrea/handle/123456789/2555/memoire.pdf

[36] Adrien Haccoun. Comparaison de méthodes de classifications. https://www.lri.fr/~antoine/Courses/Master-ISI/ISI-10/Projets_2012/Projet_DM.pdf.

[37] A, D. (2015). Classi cation Arbres de décision.

[38] Borzsonyi S., Kossman D., et Stocher K. , 'The skyline opretor', In ICDE, 2001 pp. 421-430.

[39] Alem, A., Dahmani, Y., & Mebarek, B. (2019). Skyline computation for improving naïve Bayesian classifier in intrusion detection system. Ingenierie Des Systemes d'Information, 24(5), 513–518. https://doi.org/10.18280/isi.240508

[40] Bouazza.A, Chaabi .A., Détection des attaques de routage dans l'Internet des Objets, mémoire de master informatique ,Université Ibn Khaldoun Tiaret .septembre 2020 .

[41] (Yavuz et al., 2018) Yavuz, F. Y., Ünal, D., & Gül, E. (2018). Deep learning for detection of routing attacks in the internet of things. International Journal of Computational Intelligence Systems, 12(1), 39–58.

[42] Sharma, M., Elmiligi, H., Gebali, F., & Verma, A. (2019). Simulating Attacks for RPL and Generating Multi-class Dataset for Supervised Machine Learning. 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2019, 20–26. https://doi.org/10.1109/IEMCON.2019.8936142.

[43] Witten, I.H., Frank, E. (2000). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementation. Morgan Kaufmann Publishers, San Francisco, California, 2000.