RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université de SAIDA Dr Moulay Taher

Faculté de Technologie

Département d'informatique

Mémoire de master

Option:Sécurité Informatique et Cryptographie

L'utilisation de l'apprentissage automatique pour la detection des botnets dans L'IOT

Présenté par:

LALA Amina

Encadré par :

Dr BENYAHIA Kadda

Année universitaire 2020/2021

Résumé

Actuellement, et avec l’augmentation énorme des objets utilisant l’internet, les botnets constituent une menace majeure pour ses objets. Un Botnet est un réseau d'ordinateurs compromis sous l'influence du code Bot (malware). Dans ce mémoire nous discutons une technique de détection des botnets basé sur les algorithmes d’apprentissage numérique. Nous avons appliqué algorithmes de classification sur un ensemble de données capturés. Enfin, nous discutons les performances de chaque algorithme afin de choisir le meilleur algorithme permettant la détection des trafics botnets qui est dans notre cas Random-forest.

Abstract

Currently, and with the huge increase in objects using the internet, botnets pose a major threat to its objects. A Botnet is a network of computers compromised under the influence of Bot code (malware). In this thesis we discuss a botnet detection technique based on digital learning algorithms. We applied classification algorithms on a captured data set. Finally, we discuss the performance of each algorithm in order to choose the best algorithm allowing the detection of botnet traffic which in our case is Random-forest.

ملخص

في الوقت الحالي ، ومع الزيادة الهائلة في الكائنات التي تستخدم الإنترنت ، تشكل شبكات الروبوت تهديدًا كبيرًا لأهدافها. إن الروبوتات عبارة عن شبكة من أجهزة الكمبيوتر التي تم اختراقها تحت تأثير رمز بوت(برنامج ضار). في هذه الرسالة ، نناقش تقنية اكتشاف الروبوتات بناءً على خوارزميات التعلم الرقمي. قمنا بتطبيق خوارزميات التصنيف على مجموعة البيانات الملتقطة. أخيرًا ، نناقش أداء كل خوارزمية من أجل اختيار أفضل خوارزمية تسمح باكتشاف حركة مرور الروبوتات والتي هي في حالتنا الغابة العشوائية.

Remerciements

En tout premier lieu, jeremercions le bon **Dieu**, tout puissant, de m'avoir donné la force pour survivre, ainsi que l'audace pour dépasser toutes les difficultés.

La première personne que je tiens à remercier est mon **Encadreur Mr BenyahiaKadda**pour l'orientation, la confiance et la patience qui ont constitué un apport considérable sans lequel ce travail n'aurait pas pu être mené au bon port. Qu'il trouve dans ce travail un hommage vivant à sa haute personnalité.

Mes remerciements s'adressent aussi à tous les **enseignants** qui m'onenseigné et qui par leurs compétences m'on soutenu dans la poursuite de mes études, ainsi qu'à tous les enseignants du département d'informatique

Je remercie mes très chers **parents**, qui ont toujours été là pour moi, « Vous avez tout sacrifié pour vos enfants n'épargnant ni santé ni efforts. Vous m'avez donné un magnifique modèle de labeur et de persévérance. Je suis redevable d'une éducation dont je suis fière ».

Enfin, je remercie **Tous** ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail.

Dedicace

Je dédie ce modeste travail

A ma très chère mère

Affable, honorable, aimable, Tu représentes pour moi le

Symbole de la bonté par excellence, la source de tendresse et

L'exemple du dévouement qui n'a pas cessé de m'encourager etdeprier pour moi.

Aucune dédicace ne saurait être assez éloquente pour

Exprimer ce que tu mérites pour tous les sacrifices que tu n'ascessé de me donner depuis ma naissance, durant mon enfanceet même à l'âge adulte.

A la mémoire de mon Père

Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eupour vous.

Rien au monde ne vaut les efforts fournis jour etnuit pour mon éducation et mon bien être.

Ce travail est le fruit de tes sacrifices que tu as

Consentis pour mon éducation et ma formation (Allah yerhmak).

A Mes très chers frères et sœurs ainsi que toute la famille.

Table des matières

Chapitre III: Machine Learning

Chapitre IV : Expérimentation & Résultats

Liste de figures

Liste des tableaux

Introduction générale

Introduction générale :

L'Internet des objets (IoT) ne concerne pas seulement les appareils connectés, il concerne les informations que ces appareils collectent et les informations puissantes et immédiates qui peuvent être obtenues à partir de ces informations. Ces informations peuvent être utilisées pour transformer votre entreprise et réduire les coûts grâce à des améliorations telles que la réduction des déchets de matériaux, la rationalisation des processus opérationnels et mécaniques ou l'expansion dans de nouveaux secteurs d'activité qui ne sont rendus possibles qu'avec des données fiables en temps réel [2]. Il existe une grande quantité de flux de données entre les appareils connectés, et donc la sécurité est devenue la principale préoccupation. Étant donné que ces derniers connectent des objets à Internet, et que ces objets peuvent communiquer entre eux sans aucune intervention manuelle, un certain nombre de nouvelles vulnérabilités sont découvertes chaque jour avec le l'introduction de nouveaux schémas d'attaque.Botnet est l'une des menaces les plus importantes contre le cyber Sécurité.

Le botnet est un exemple d'utilisation de bonnes technologies pour de mauvaises intentions. Un botnet n'est rien de plus qu'une chaîne d'ordinateurs connectés coordonnés ensemble pour effectuer une tâche. Cela peut être le maintien d'une salle de discussion ou la prise de contrôle de votre ordinateur. Les botnets ne sont qu'un des nombreux dangers présents sur Internet.

Les techniques d'apprentissage automatique (machine learning) comme la classification et le clustering, qui intègrent des algorithmes d'induction qui explorent les données afin de découvrir des modèles cachés et de développer des modèles prédictifs, se sont avérés être efficace pour relever les défis susmentionnés en matière de cybersécurité. Le volume de données traitant à la fois des activités du réseau et de l'hôte est si important et le développement de modèles prédictifs est essentiel, ce qui rend les techniques d'apprentissage automatique adaptées à la détection des botnets.

Dans le présent travail, nous avons appliqué l'apprentissage automatique pour la détection des botnets dans les IoT , nous appliquons des différents algorithmes d'apprentissage automatique (machine Learning ML) sur un ensemble de données qui représente des captures des trafics normal et botnets Pour tirer le meilleur modèle qui aide à atténuer les trafics botnets.

Ce mémoire est structuré en quatre chapitres :

- ✓**Chapitre 1 :** Présente l'internet des objets, citer ses fonctionnalités, ses différents domaines d'applications et expliquer ses caractéristiques.
- ✓**Chapitre 2 :** est consacré à la présentation des bot-net, leurs architectures, les types d'attaque et expliquer son fonctionnement
- ✓**Chapitre 3 :** est un état des lieux autour de l'apprentissage automatique (machine Learning ML), les types d'apprentissage et ces algorithmes supervisé et leurs avantages et inconvénients
- ✓**Chapitre 4 :** Dans lequel nous avons présenté les expérimentations réalisées et les résultats obtenus.

Chapitre I : Généralité sur IOT

I.1 Introduction :

L'internet des objets consiste de manière simplifiée à connecter des objets. En quelque sorte il s'agit de l'extension de l'Internet au monde réel des objets qui nous entourent. Elle apporte un grand bénéfice : simplification des taches quotidiennes, meilleur gestion de l'énergie, facilite la vie à la personne handicapée, amélioration des suivis de santé… Dans ce chapitre, nous définissons de manière générale l'internet des objets.

I.2 Définition de l'IoT :

Une recommandation *Présentation générale de l'Internet des objets* (ITU-T Y.2060), juin 2012, § 3.2.2 Définition internationale (par l'Union internationale des télécommunications[2] définit l'Internet des objets comme une « infrastructure mondiale pour la société de l'information, qui permet de disposer de services évolués en interconnectant des objets (physiques ou virtuels) grâce aux technologies de l'information et de la communication interopérables existantes ou en évolution » [3] Pour l'Union, en exploitant les capacités d'identification, de saisie de données, de traitement et de communication, l'IdO tire pleinement parti des objets pour offrir des services à toutes sortes d'applications, tout en garantissant le respect des exigences de sécurité et de confidentialité. Elle note enfin que, dans une optique plus large, l'IdO peut être considéré comme un concept ayant des répercussions sur les technologies et la société [4]

L'IdO est donc « un réseau de 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 » [5]

D'autres définitions insistent sur les aspects techniques de l'IdO (« des objets ayant des identités et des personnalités virtuelles, opérant dans des espaces intelligents et utilisant des interfaces intelligentes pour se connecter et communiquer au sein de contextes d'usages variés »[6], d'autres portent sur les usages et les fonctionnalités (« la convergence des identifiants numériques »[7] notant qu'il devient possible d'identifier de manière unifiée des éléments d'information numérique (adresses) et des éléments physiques (une palette dans un entrepôt, ou un animal dans un troupeau).

I.3 Historique de l'internet des objets :

L'Internet des objets est apparu dans le cadre d'une tendance [8] lourde, issue de la mécanisation et la standardisation, appliquée à l'automatisation du traitement du document et

de l'information sur support matériel puis numérique (dont au service de la production et recherche documentaire). Apparu aux États-Unis dès 1982[9], il s'est rapidement diffusé avec la mondialisation, aboutissant à connecter des machines à des serveurs capables de les superviser (ces machines étant notamment des ordinateurs mis en réseau dans ce que certains ont nommé l'« Internet des machines ») [10]. Peu à peu des objets ont été modifiés (avec des puces RFID par exemple) ou conçus pour « parler le protocole IP », devenant des « objets connectés », reliés à des serveurs centralisés ou capables de communiquer entre eux ou avec des réseaux de serveurs et divers acteurs, d'une manière de moins en moins centralisée.

Ses enjeux diffèrent selon les pays ou les régions du monde, et selon les acteurs et « leurs intérêts parfois divergents » [11]. Ce mouvement s'est accompagné d'une croissance et d'une complexification des systèmes de sécurité (pare-feux, mots de passe, etc.).

Il est parfois suggéré que l'objet deviendra un acteur autonome de l'Internet, capable de percevoir, d'analyser et d'agir de lui-même selon les contextes ou les processus dans lesquels il sera engagé [12]. Dans ce cas de figure, l'avènement de l'Internet des objets s'associe à celui des technologies ou des méthodes de conception logicielle liées à l'Intelligence artificielle et des sciences de la complexité. Le couple « objet physique » / « intelligence virtuelle associée », que cette dernière soit embarquée, distribuée ou hébergée dans le Cloud (cloudcomputing), y est alors mentionné sous l'appellation de « cyberobjet », ou encore « d'avatar digital » [13], concept repris par la suite dans la notion de « Jumeau numérique ». Les cyberobjets sont des acteurs potentiels des chaînes de valeurs qui agissent sous le contrôle des opérationnels ou en partenariat avec eux. En accédant ainsi au statut d'assistants, de conseillers, de décideurs ou encore d'organisateurs (selon les cas), ils deviennent de véritables agents économiques [14] et contribuent à la mutation des modèles économiques ou de gestion existants.

Deux enjeux récurrents sont la protection de la vie privée (« privacy ») et de la régulation d'une part [15] et la gouvernance de cet Internet d'autre part, de plus en plus ubiquitaire et multiforme, quand il n'y a plus d'interface unique [16, 17]. En France, en 2015 et en 2016, le forum international IoTPlanet s'est déroulé au mois de novembre à Grenoble afin de faire le point sur l'évolution technologique des objets connectés [18]

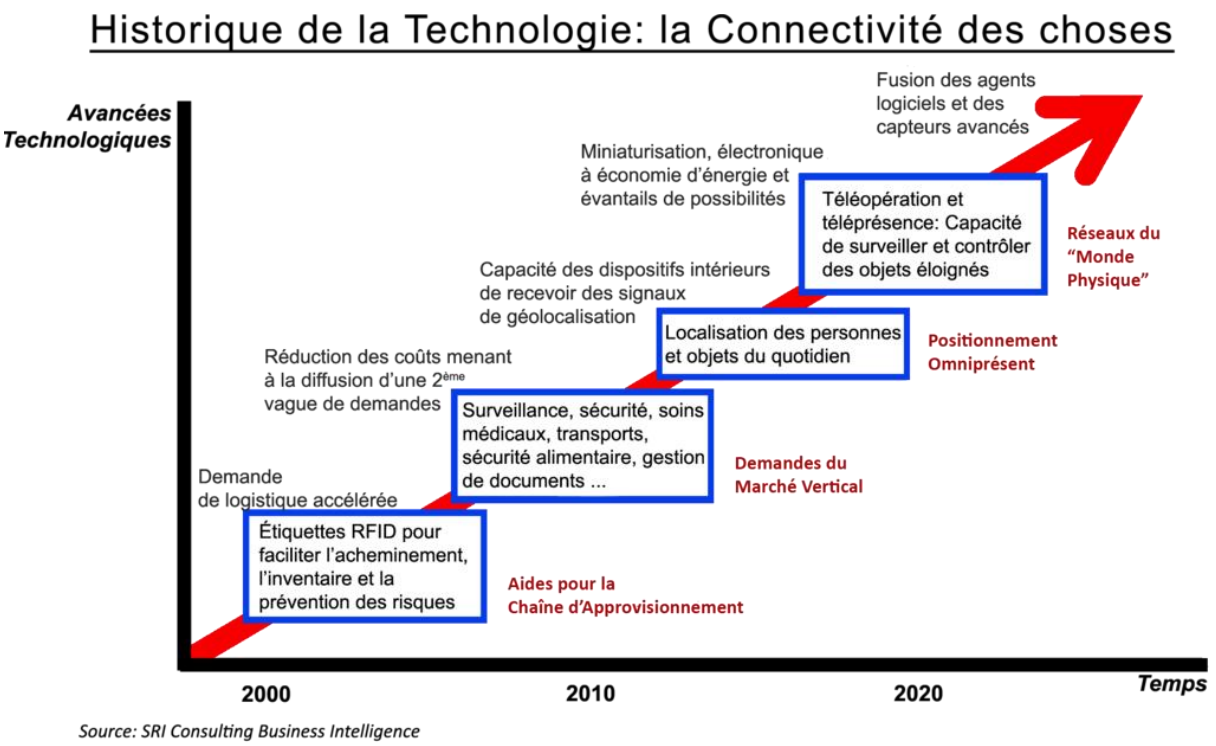


Figure I .1 : historique de la technologie la connectivité des choses

I.4 Caractéristique de l'Internet des objets :

2.2 Caractéristiques d'un objet connecté Généralement, un objet connecté est caractérisé par [19]: Identité : pour que les objets soient gérables il est essentiel que chaque objet connecté possède une identité unique qu"il lui propre et qui le distingue des autres objets du système.

I.4.1 Interactivité :

Les progrès technologiques ont permis de connecter une grande variétéd"objets et de dispositifs. Un objet n"a pas besoin d"être connecté à un réseau à tout moment. Pour des objets dits passifs tels que des livres ou des DVD, des étiquettes RFID doivent seulement être en mesure de signaler leur présence, de temps en temps, comme le moment de quitter le magasin. Programmable: l"objet connecté doit être programmé et piloté à distance via un ordinateur, une tablette ou un Smartphone.

I.4.2 Sensibilité :

Un objet a la capacité de percevoir son environnement et peut collecter ou transmettre des informations à celui-ci. Il peut ainsi avoir des capteurs signalant les niveaux de température, d"humidité, de vibrations, d"emplacement ou debruit.

I.4.3 Autonomie :

Cette caractéristique est, peut-être, la caractéristique la plus importante pourl"objet connecté. On désigne par cette caractéristique la capacité de l"objetd"agir sans l"interventiond"un tiers.

En d‟autres termes, les objets doivent pouvoir être traités et surveillés individuellement, généralement depuis un point éloigné, et doivent fonctionner indépendamment d‟une télécommande, c.-à-d. que chaque objet devient responsable de lui-même.

I.5 Architecture de l'internet des objets :

L'un des principaux défis à relever dans le domaine technologique pour favoriser le déploiement des systèmes IoT est de définir une architecture de référence prenant en charge les fonctionnalités actuelles etextensions futures. Actuellement, il n'existe pas d'architecture de référence unique, et en créer une s'avère trèscompliqué malgré de nombreux efforts de standardisation. Le principal problème réside dans la naturefragmentation des applications possibles, dont chacune dépend de nombreux très souvent différent variables et spécifications de conception.

Ceproblème doit être ajouté à chaque fournisseur tendance à proposer sa plate-forme pour des applications similaires [20–22]. Dans la figure 1, il est possible de voir certaines des architectures IoT les plus courantes utilisées

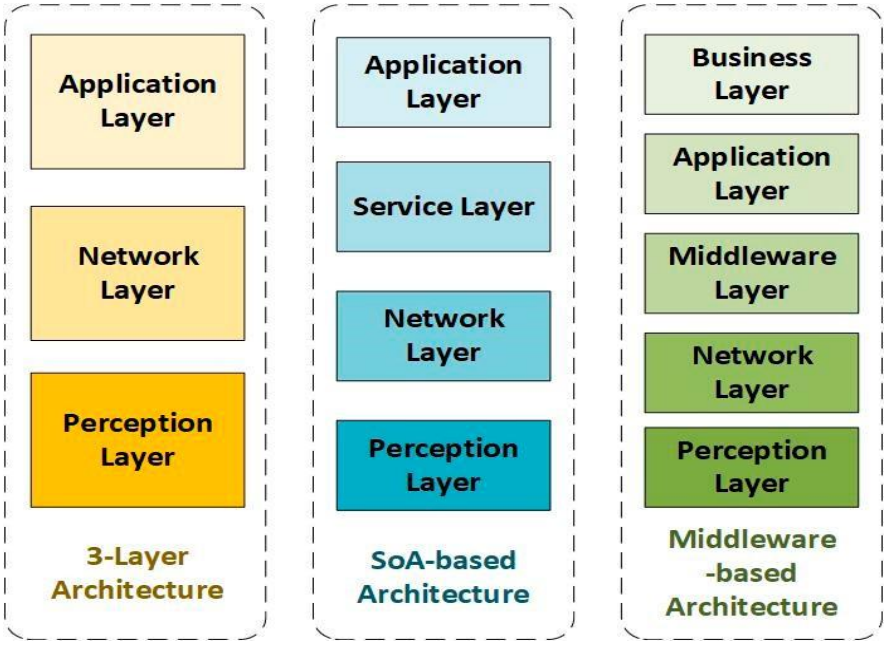


Figure I.2 : architectureIoT les plus courantes

I.5.1 Architectures à trois couches :

Une architecture générique de haut niveau composée de trois couches a été introduite dans la Littérature [dix] :

Perception, qui représente la couche physique des objets et regroupe toutes les caractéristiques Réseau, qui représente la couche de communication responsable de la

transmissiondes données à la couche applicative via diverses technologies et protocoles ;
Application, qui représente la couche applicative dans laquelle le logiciel offrant unservice spécifique est effectivement mis en œuvre.

I.5.1.1 Couche de perception :

La couche de perception représente le niveau physique des objets et interagit avec la surface arrondir l'environnement en collectant et en traitant les informations. Ce niveau comprend des objets qui peu, être capable d'interagir avec le monde extérieur et être équipé en informatiquecapacités, deviennent dans un certain sens « intelligents » ou « intelligents », où intelligent fait référence à laaspects technologiques (les technologies intelligentes utilisées), tandis qu'intelligent fait référence auxaspects (auto-identification, auto-diagnostic, auto-test, etc.) du capteur [23, 24].

Ces objets intelligents, qui sont les blocs fondamentaux sur lesquels repose l'IoT, peuvent être des objets d'usage courant (un réfrigérateur, une télévision, une voiture, etc.) ou de simples appareilséquipés de capteurs et de capacités de calcul. En général, les objets intelligents sont équipés avec les propriétés essentielles suivantes [25 , 26] :

Communication : les objets peuvent se connecter entre eux et à des ressources sur Internet pour utiliser les données et les services, mettre à jour leur statut et coopérer pour atteindre des objectifs communs ;

Identification : les objets doivent être identifiés de manière unique.

Selon l'application spécifique, une ou plusieurs des propriétés suivantes peuvent s'ajoutent également :

Adressabilité : les objets peuvent être directement accessibles, c'est-à-dire adressés, pour être interrogés et/ou configuré à distance ;

Détection et actionnement : les objets peuvent collecter des informations sur le monde environnantet le manipuler à l'aide de capteurs et d'actionneurs ;

Traitement de l'information embarqué : les objets intelligents sont dotés de capacités de calcul pour traiter les résultats des capteurs et piloter les actionneurs ; Localisation : les objets connaissent leur emplacement physique ou peuvent être localisés

Interface utilisateur : les objets peuvent communiquer de manière appropriée avec les utilisateurs via des écrans ou d'autres interfaces.

I.5.1.2 Couche réseau :

La couche réseau a pour tâche de transporter les données fournies par la perception niveau à la couche d'application. Il comprend toutes les technologies et tous les protocoles qui rendent ce connexion possible et ne doit pas être confondu avec la couche réseau de l'ISO/OSI

modèle, qui achemine uniquement les données au sein du réseau le long du meilleur chemin [27].

coucheréseau.Dans cette couche, les protocoles sans fil sont particulièrement importants. Par rapport à ceux qui nécessitentcâbles, les capteurs sans fil peuvent être installés dans des environnements difficiles d'accès et nécessitent moinsmoyens matériels et humains pour l'installation. De plus, dans un réseau de capteurs sans fil,les différents nœuds peuvent être ajoutés ou supprimés facilement, et leur emplacement peut être modifiésans remettre en cause la structure de l'ensemble du réseau. Le choix d'un protocole à utiliserdépend de la taille du réseau, de la consommation électrique de chaque nœud et de la transmissionvitesse nécessaire dans une application donnée.

Dans d'autres applications, cependant, il peut être nécessaire de construire un réseau câblé. Le dernierbénéficie d'une fiabilité plus excellente et de taux de transmission plus élevés [14]. Pour donner un exemple,

il est possible de penser au réseau interne d'un véhicule qui relie les différents

Unités de contrôle (ECU) qui contrôlent les parties mécaniques de la voiture (direction, frein, etc.). Dans ce cas, il est indispensable d'avoir un réseau fiable et rapide, car des retards ou des dysfonctionnements pourrait avoir de graves conséquences pour les personnes à bord de la voiture.

I.5.1.3 Couche d'application :

La couche applicative comprend tous les logiciels nécessaires pour offrir un service spécifique. Dans ce niveau, les données des niveaux précédents sont stockés, agrégées, filtrées et traitées, et des bases de données, des logiciels d'analyse, etc., sont utilisés. À la suite de ce processus de traitement.

Les données sont mises à disposition de véritables applications IoT (smart wearable, smart car, etc.). C'estsouvent fait à l'aide de certains logiciels définis comme middleware, qui ont pour tâche de cacher la hétérogénéité des couches sous-jacentes. Certaines technologies logicielles actuellement largement utiliséespour gérer la grande quantité de données fournies par les appareils sont :

Cloud computing, où des services tels que le stockage ou le traitement de données sont fournisà partir d'un ensemble de ressources préexistantes, paramétrables et disponibles à distance sous la formede l'architecture distribuée ;

Edgecomputing, où le traitement des données est partiellement distribué sur le périphériquenœuds du réseau pour augmenter les performances des systèmes IoT.

La gestion du format des données à traiter appartient également à ce niveau.

Ceux-ci peuvent être du type [29] : à base binaire, de petite taille mais non lisible par les êtres humains ; texte, de plus grande taille mais lisible par les êtres humains.

Parmi les nombreuses plateformes commerciales utilisées pour la mise en œuvre d'applications IoT, certaines des exemples sont Amazon AWS, Microsoft Azure, Xively, Firefox WebThings Gateway, etc

I.5.2 Architecture à cinq couches :

L'architecture en trois couches définit l'idée principale de l'Internet des objets, mais elle n'est pas suffisante pour la recherche sur l'Internet des objets parce que la recherche se concentre souvent sur des aspects plus fins de l'Internet des objets. C'est pourquoi, nous avons beaucoup plus d'architectures stratifiées proposées dans la littérature. Le premier est l'architecture à cinq couches, qui comprend également les couches de traitement et d'affaires. Les cinq couches sont les couches perception, transport, traitement, application et affaires. Le rôle des couches perception et application est le même que celui de l'architecture à trois couches. Nous décrivons la fonction des trois autres couches [30].

I.5.2.1 Couche de transport : Cette couche transfère les données du capteur de la couche de perception à la couche de traitement et vice versa à travers des réseaux tels que sans fil, 3G, LAN, Bluetooth, RFID, [30].

I.5.2.2 Couche de traitement : Cette couche est également connue sous le nom de couche de middleware. Elle stocke, analyse et traite d'énormes quantités de données qui proviennent de la couche de transport. Elle peut gérer et fournir un ensemble diversifié de services aux couches inférieures. Elle utilise de nombreuses technologies comme les bases de données, l'informatique en cloud et les modules de traitement des méga-données [30].

I.5.2.3 Couche d'affaires : Cette couche gère l'ensemble du système de l'Internet des objets, y compris les applications, les modèles d'affaires et de profit, et la vie privée des utilisateurs.

Le niveau opérationnel n'est pas visé par le présent document.Par conséquent, nous n'en discutons pas davantage [30].

I.6 Technologies de communication de l'IdO :

Le concept IoT dans le monde réel peut être réalisé grâce à l'intégration de plusieurs technologies habilitantes. Dans cette section, les technologies habilitantes les plus pertinentes pour l'IoT sont présentées, en se concentrant sur les différentes couches énumérées ci-dessus. Au fond de toute architecture, nous pouvons trouver la couche de perception, qui

contient tous les appareils et objets physiques : dans cette couche, la fonction principale est d'identifier et de suivre les objets. Pour réaliser cette fonction, plusieurs technologies peuvent être mises en œuvre :

I.6.1 Identification par radiofréquence (RFID) :

Radio Frequency Identification : méthode utilisée pour stocker et récupérer des données à distance en utilisant les Tags RFID. Ces Tags, qui peuvent être collées ou incorporées dans des produits, et qui sont composées d'une antenne et d'une puce électronique, réagissent aux ondes radio et transmettent des informations à distance. Cette technologie a été développé dans l'objectifs de remplacer les codes-barres [31]

I.6.2 IEEE 802.15.4 wifi :

Le wifi caractérisé par La norme *IEEE 802.11* qui est un standard international son débit maximal peut varier de 6 Mbits/s a 54 Mbits/s en intérieur la fréquence varie selon les différentes normes de 2,4GHz et 5GHz L'omniprésence du Wifi dans l'environnement de la maison dans les réseaux locaux. Il nécessite peu d'explication supplémentaire, Actuellement la norme Wifi le plus couramment utilisé dans les foyers et de nombreuses entreprises est 802.11n, qui offre un débit sérieux dans la gamme de centaines de mégabits par seconde, ce qui est très bien pour les transferts de fichiers, mais peut-être trop consommateur d'énergie pour de nombreuses applications de l'IdO. [32]

I.6.3 ZensysWave (Z-Wave) :

Est une technologie de faible puissance de communication RF qui est principalement conçu pour l'automatisation de la maison pour les produits tels que les contrôleurs de la lampe et les capteurs parmi beaucoup d'autres. Optimisé pour une communication fiable et à faible latence des petits paquets de données avec des débits de données jusqu'à 100Kbit /s, il fonctionne dans la bande sous-1GHz et est imperméable aux interférences provoquées par WiFi et d'autres technologies sans fil dans la gamme de 2,4 GHz, tels que Bluetooth ou ZigBee. [33]

I.6.4 Bluetooth :

Bluetooth est une solution de connectivité sans fil dominante de courte portée permet d'obtenir des débits de l'ordre de 1 Mbps, utilise les ondes radio de bande de fréquence de 2.4 GHz. Il a une pénétration universelle dans l'espace de l'appareil mobile et est largement intégré dans les ordinateurs personnels, les Smartphone, et les accessoire grand public, l'évolution de la norme Bluetooth intelligente et Bluetooth intelligente Ready et maintenant Bluetooth 5.0, et l'ajout de TCP / IP et de maillage capacités de réseautage, [34]

I.6.5 LoWPAN :

Une clé IP (Internet Protocol) à base est 6LoWPAN (IPv6 faible puissance sans fil Personal Area Network), 6LoWPAN est un protocole réseau qui définit les mécanismes d'encapsulation et de compression d'en-tête. La norme a la liberté de la bande de fréquence et de la couche physique et peut également être utilisé sur plusieurs plates-formes de communication, y compris Ethernet, Wi-Fi, 802.15.4 et sous-1GHz ISM. Un attribut clé est la pile IPv6 (Internet Protocol version 6), IPv6 est le successeur de l'IPv4 et offre environ 5 x 10^28 adresses pour chaque personne dans le monde, permettant à tout objet ou dispositif embarqué dans le monde d'avoir sa propre adresse IP unique et se connecter àInternet.
Spécialement conçu pour la maison ou l'automatisation des bâtiments. [35]

I.6.7 Near-field communication (NFC) :

(Near Field Communication) est une technologie qui permet des interactions à deux voies simples et sûres entre des appareils électroniques, et surtout applicable pour les Smartphones, permettant aux consommateurs d'effectuer des transactions de paiement sans contact, le contenu numérique d'accès et de connecter des appareils électroniques. Essentiellement, il étend la capacité de la technologie de carte sans contact et permet aux périphériques de partager des informations à une distance qui est inférieure à 10cm.[36]

I.6.8 Ultra-wide band (UWB):

La technologie de communication UWB est conçue pour prendre en charge les communications dans les zones de couverture à faible distance, ce qui est similaire à la technologie NFC qui utilise une faible énergie.Ce pendant, une large bande passante est utilisée pour les applications permettant de connecter des capteurs pour la communication. Elle est capable d'une largeur de bande maximale de 500 MHz.

I.6 .9 Machine to Machine (M2M):

M2M désigne les communications entre ordinateurs, processeurs intégrés, capteurs intelligents, actionneurs ou dispositifs mobiles. Il existe au total cinq composantes de base de la technique de communication M2M : détection, accès hétérogène, traitement de l'information, applications et services. M2M est une structure en cinq parties qui comprend les parties suivantes [37] :

I.6.9.1 Dispositif M2M : Dispositif capable de répondre aux demandes de données contenues dans ce dispositif [37].

I.6.9.2 Réseau M2M (domaine des appareils) : Fournir une connectivité entre les appareils M2M et les passerelles M2M [37].

I.6.9.3 Passerelle M2M : Utiliser les capacités M2M pour assurer l'interconnexion des dispositifs M2M au réseau de communication [37].

I.6.9.4 Réseaux de communication M2M (domaine de réseau) : Communications entre la passerelle M2M et l'application M2M [37].

I.6.9.5 Applications M2M : Contient la couche de middleware où les données passent par divers services d'application et sont utilisées par les moteurs de traitement métier spécifiques [37].

Services d'application et sont utilisées par les moteurs de traitement métier spécifiques [37].

I.7 La fonctionnalité de l'internet des objets :

L'écosystèmeIoT est assez complexe, car il intègre plusieurs technologies et domaines de compétences. Un système IoT englobe, généralement, à la fois du hardware, des protocoles de communication, du software, du cloud et du mobile. Ainsi, un projet IoT nécessite d'avoir une équipe pluridisciplinaire On peut décomposer un système IoT en 4 fonctionnalités distinctes comme le montre la figure ci-dessous :

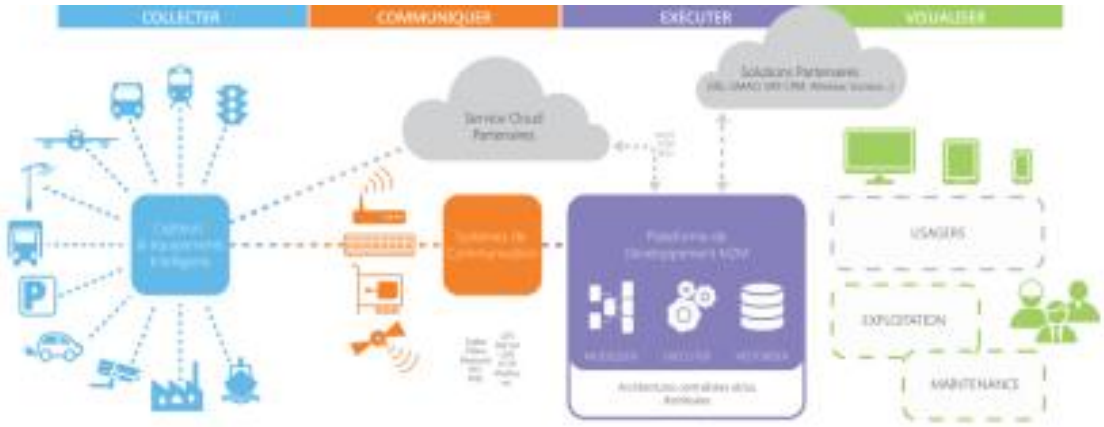


Figure I.3 : les fonctionnalités d'un écosystème IOT

I.7.1 Collecter / Actionner :

A cette étape, on est au niveau de l'objet connecté. On parle de capteurs qui permettent de faire des mesures de l'environnement physique (ex : température, humidité, bruit) et des actionneurs qui peuvent agir sur l'environnement (ex: des moteurs pour fermer ou ouvrir une porte). Certains objets peuvent être dotés de capacités électroniques, informatiques et réseaux qui leur permettent de se connecter directement au réseau Internet. Mais généralement, ayant des contraintes matérielles et logicielles (autonomie limité, capacité de traitement limité, pas de stack réseau, etc), les objets implémentent des protocoles de communication à basse

énergie / bas débit et communiquent avec le réseau internet à travers une passerelle « gateway ». Un exemple de cette Gateway est notre téléphone portable qui permet de collecter les données de ma smartwatch. [38]

I.7.2 Communiquer :

À cette étape que se passe l'envoi des données du LAN vers le Cloud. Et on peut distinguer deux modèles de protocoles pour transporter la donnée : Le modèle Publish / Subscribe avec des protocoles de type MQTT et le modèle REST avec des protocoles comme HTTP ou encore CoAP. [38]

I.7.3 Exécuter :

C'est l'étape de stockage et de traitement de la donnée. À cette étape on parle souvent de « Plate-forme IoT » qui est souvent une solution cloud capable de connecter plusieurs objets connectés, stocker leurs données, les traiter, les analyser et les exposer à travers différentes applications. Les plateformes IoT permettent aussi de faire communiquer de objets hétérogènes. Ces plateformes se multiplient de nos jours (Amazon, Google, Microsoft, etc.) et on parle même de « guerre des plateformes IoT» .[38]

I.7.4 Visualiser :

C'est l'étape qui permet d'exposer les services des objets connectés à travers différentes applications dédiées. Un utilisateur, à travers une application mobile, peut par exemple communiquer avec ses objets en consultant leurs données ou en envoyant des actions vers ses objets. [38]

I.8 Le Domaine D'Application

Nous avons recensé de nombreuses applications IdO et nous les avons classées

I.8.1 La domotique :

Cette catégorie regroupe les appareils de contrôle à distance : allumer et éteindre les appareils à distance pour éviter les accidents et économiser de l'énergie, l'utilisation de l'énergie et de l'eau : surveillance de la consommation d'énergie et d'eau pour obtenir des conseils sur la façon d'économiser les coûts et les ressources, L'art et préservation des biens : suivi de l'état de conservation à l'intérieur des musées et des entrepôts d'art, et les systèmes de détection d'intrusion : détection des ouvertures de portes, de fenêtres et des violations dans le but d'empêcher les intrusions. L'éclairage intelligent attire une attention croissante de la communauté de recherche [40] HomeKit [39] est un framework conçu par Apple permettant aux utilisateurs de configurer, communiquer et contrôler des appareils domestiques intelligents. Les utilisateurs peuvent effectuer des actions automatiques dans la maison au

moyen d'une simple dictée vocale.

I.8.2 Environnement intelligent :

Cette catégorie regroupe la détection précoce des tremblements de terre : contrôle distribué dans des endroits spécifiques de tremblements, les glissements de terrain et la prévention des avalanches : surveillance de l'humidité du sol, des vibrations et de la densité de la terre pour détecter les tendances dangereuses dans les conditions du terrain, la surveillance du niveau de neige : mesure de niveau de neige pour connaître en temps réel la qualité des pistes de ski et permettre la sécurité des avalanches, la détection des incendies de forêt : surveillance des gaz de combustion et des conditions d'incendie pour définir les zones d'alerte, et la pollution de l'air : contrôle des émissions de CO2 des usines, de la pollution émise par les voitures et des gaz toxiques. Insightrobotics [41] détecte les incendies de forêt en fusionnant les informations collectées par des caméras en réseau et différents types de capteurs (vent, température, etc.).

I.8.3 Le transport et logistique :

Cette catégorie regroupe la détection d'incompatibilité de stockage : émissions de conteneurs stockant des produits inflammables fermés à d'autres contenant des matières explosives, le suivi de flotte : contrôle du suivi des itinéraires pour les marchandises sensibles comme les bijoux, les médicaments ou les marchandises dangereuses, l'emplacement des articles : recherche d'éléments individuels dans de grandes surfaces comme les entrepôts ou les ports et la qualité des conditions d'expédition : surveillance, à des fins d'assurance, des vibrations, des coups, des ouvertures de conteneurs ou de leur entretien. HiKoB [42] fournit une gestion et des informations en temps réel sur les conditions de circulation et des services pour le transport de marchandises et la logistique. HiKoB collecte des mesures en temps réel telles que les températures extérieures actuelles, l'humidité, les points de rosée et de givre, les gradients de température à partir de capteurs déployés sur les routes Alltrafficsolutions [43] collecte des données sur le trafic routier au moyen de capteurs et les visualise sur des cartes afin de fournir aux conducteurs des informations actualisées. Il prend en compte les modifications des panneaux de signalisation numériques, les panneaux à message variables ou les panneaux de limitation de vitesse.

I.8.4 Cybersanté :

Cette catégorie regroupe le rayonnement ultraviolet : mesure des rayons UV pour prévenir les personnes en cas de forte exposition, les soins aux sportifs : surveillance des signes vitaux dans les centres et les champs de haute performance, le suivi des personnes seules : assistance aux personnes âgées ou handicapées vivantes en autonomie, la surveillance des patients : suivi des conditions des patients à l'intérieur des hôpitaux et dans la maison de retraite, et les

réfrigérateurs médicaux : contrôle des conditions à l'intérieur des congélateurs stockant les vaccins, les médicaments et les éléments organiques. En général, les systèmes collectent les données vitales des patients via un réseau de capteurs connectés aux dispositifs médicaux et garantissent l'accès ubiquitaire ou le partage de données médicales comme l'ElectronicHealthcare Records (EHR) [44].

I.8.5 Contrôle industriel :

Cette catégorie regroupe la mesure de la qualité de l'air intérieur : surveillance des niveaux de gaz toxiques et d'oxygène à l'intérieur des usines chimiques pour assurer la sécurité des travailleurs et des biens, la surveillance de la température : contrôle de la température à l'intérieur des réfrigérateurs industriels et médicaux avec des marchandises sensibles, l'auto-diagnostique du véhicule : collecte d'informations sur le bus interne du véhicule afin d'envoyer des alarmes en temps réel aux urgences ou fournir des conseils aux conducteurs, et la localisation à l'intérieur : emplacement intérieur des ressources en utilisant des étiquettes actives et passives. Yanzi [45] est une solution permettant de surveiller, d'entretenir et de gérer les ascenseurs et les systèmes de chauffage. Les informations sont récupérées via des capteurs vidéo, de mouvement, de température et de lumière.

I.8.6 Agriculture intelligente :

Cette catégorie regroupe le compost : contrôle de l'humidité et des niveaux de température dans le foin, la paille, etc. pour prévenir les champignons et autres contaminants microbiens, les stations météorologiques : étude des conditions météorologiques dans les champs pour prévoir la formation de glace, la pluie, la sécheresse, la neige ou les changements de vent, l'amélioration de la qualité du vin : surveiller l'humidité du sol et le diamètre du tronc dans les vignes pour contrôler la quantité de sucre dans la vigne et sa santé, les cours de golf : l'irrigation sélective dans les zones sèches pour réduire les ressources en eau nécessaires, les serres : contrôler les conditions microclimatiques pour maximiser la production de fruits et légumes et sa qualité et l'hydroponique : contrôler l'état des plantes cultivées dans l'eau pour obtenir les cultures les plus efficaces. L'agriculture devient de plus en plus complexe et interconnectée. OnFarm [46] facilite sa gestion. Les informations contextuelles comme la cartographie, la localisation, l'humidité du sol, la télémétrie et la météo sont utilisées pour une prise de décision efficace en temps réel.

I.8.7 Sécurité et Urgences :

Cette catégorie regroupe les mesures de niveaux de rayonnement : mesure distribuée des niveaux de rayonnement dans les environs des centrales nucléaires pour générer des alertes de fuite, le contrôle d'accès périmétrique : contrôle d'accès aux zones restreintes et détection des

personnes dans les zones non autorisées, les gaz explosifs et dangereux : détection des niveaux de gaz et des fuites dans les environnements industriels, les environnements des usines chimiques et l'intérieur des mines, et la présence liquide : détection de liquides dans les centres de données, les entrepôts et les terrains sensibles afin de prévenir les pannes et la corrosion. Aircasting [47] est une plate-forme d'enregistrement, de cartographie et de partage de données relatives à la santé et à l'environnement obtenues à l'aide de smartphones et de dispositifs de surveillance. Les informations recueillies incluent les concentrations de monoxyde de carbone (CO) et de dioxyde d'azote (NO2), la température, les niveaux sonores, l'humidité, la fréquence cardiaque et respiratoire et le niveau d'activité.

Chapitre II: Les botnets

II.1 introduction :

Un réseau est un ensemble d'appareils connectés sur Internet, et
le nombre total de ces appareils augmente chaque jour. Il ne fait aucun doute que
les réseaux d'institutions financières et commerciales sont continuellement soumis à des risques de sécurité, ce qui non seulement coûte des milliards de dollars en dommages et en réparation, mais a également un effet négatif sur leur réputation. Le nombre croissant d'utilisateurs affectés par des logiciels malveillants devient un problème critique. Les botnets sont devenus la principale préoccupation, car ils constituent l'une des plus grandes menaces pour les systèmes de sécurité. Leur popularité vient de leur capacité à contrôler les ordinateurs centraux d'entreprise en infiltrant tout connecté Internet appareil qui utilise un
numérique vidéo enregistreur (DVR) [8].

Un botnet peut être défini comme un réseau de périphériques hôtes compromis qui sont utilisé s pour mener des activités malveillantes. Les ordinateurs de bureau, les smartphones,
les ordinateurs portables et les tablettes sont des exemples de tels périphériques hôtes. Un botnet se compose de trois éléments : un attaquant appelé botmaster, un serveur de commande et de contrôle (C&C) et une machine infectée appelée bot. Le botmaster a besoin d'un canal C&C pour commander les bots et coordonner les attaques malveillantes. Des
exemplede canaux C&C sont IRC, HTTP et P2P. Selon la communication des protocoles, C & C canaux peut être centralisé ou décentralisé. Les
robots sont utilisés pour envoyer des attaques par déni de service distribué
(DDoS) , des attaques de phishing , des courriers indésirables et
d' autres formes d' attaques malveillantes [48].

Le WannaCry est l'un des exemples les plus connus d'attaque de ransomware. Il a nui à des entreprises du monde entier au printemps 2017. Il s'agit d'un type de logiciel malveillant, également appelé malware, utilisé par les cybercriminels pour extorquer de l'argent aux entreprises qu'ils ciblent. Pendant ce temps, ils ont attaqué plus de 200
000 ordinateurs dans plus de 150 pays. Cela comprenait le National Health Service
du Royaume - Uni . Tout en tout, il a coûté au Royaume - Uni £ 92 millions et accourut coûts globaux pour la modique somme de £ 6 milliards d' euros [49].Au cours
de la dernière décennie, botnet étant une importante persistante menace sur l' Internet a reçu pl us de recherche mise au point et un grand nombre d' attention, comme
en témoignent par examen des documents et plusieurs enquêtes portant sur les techniques de détection de botnets. L'enquête réalisée en 2015 [50], permet de comparer les recherches

existantes. C'est la première enquête présente les méthodes de prévention et de défense sur Botnet. Il n'y a que trois documents d'enquête sur Botnet détection basés sur le trafic DNS analyse présentée par [51] en 2015, [52] en 2017, et [53] en 2019. Ils ont résumé les actuelles techniques en botnet détection
des méthodes qui sont basées sur DNS trafic analyse et techniques pour atténuer la menace des botnets. Alors que l'enquête de 2020 [54], se concentre sur les nombreuses approches et compare les différentes méthodes utilisées dans un passé récent dans la détection générale des botnets et la détection des IoT-Bot. Cependant, l'enquête de 2020 [55]
a abordé les concepts impliqués dans la défense
d' une attaque DDoS , des méthodes traditionnelles aux méthodes spécifiques à l' IoT .
Une seule enquête sur la détection de botnets dans le Software Defined Networking (SDN) a été publiée en 2018 [56], et elle n'est ni exhaustive ni ne prend en considération
divers paramètres essentiels pour une comparaison efficace. Il contient même quelques articles comme références qui ne suffisent pas à donner un aperçu complet.

II.2 Définition du malware :

Un logiciel malveillant ou maliciel, aussi dénommé logiciel nuisible ou programme malveillant ou pourriciel (de l'anglais malware [ˈmælwɛə˞][57], est un programme développé dans le but de nuire à un système informatique, sans le consentement de l'utilisateur dont l'ordinateur est infecté. Il existe plusieurs méthodes utilisées par les pirates pour infecter un ordinateur, comme le phising (hameçonnage par e-mail) ou le téléchargement automatique d'un fichier par exemple.
De nos jours, le terme « virus » est souvent employé, à tort, pour désigner toutes sortes de logiciels malveillants. En effet, les maliciels englobent les virus, les vers, les chevaux de Troie, ainsi que d'autres menaces. La catégorie des virus informatiques, qui a longtemps été la plus répandue, a cédé sa place aux chevaux de Troie en 2005.

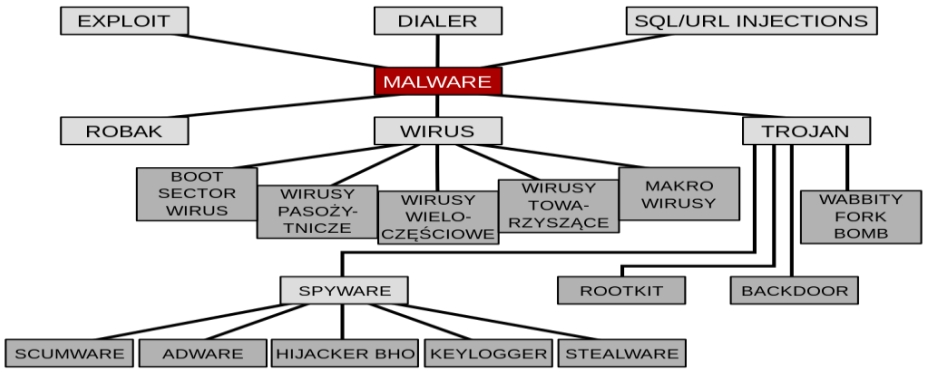


Figure II .1 : Différents types de logiciels malveillants

II.3 Définition des botnets :

Un botnet est une collection d'ordinateurs d'utilisateurs connectés à l'Internet (bots) infectés par un logiciel malveillant qui permet à ces ordinateurs d'être contrôlés à distance par un opérateur (appelé parfois bot herder) par l'intermédiaire d'un serveur de Commande-et-contrôle (C&C) pour exécuter certaines tâches comme voler de l'information ou lancer des attaques contre d'autres ordinateurs. Le logiciel malveillant des botnets est conçu pour donner à ses opérateurs le contrôle sur de nombreux ordinateurs en même temps. Ceci permet aux opérateurs de botnets d'utiliser des ressources informatiques et de bande passante à travers de nombreux réseaux pour des activités malveillantes. Historiquement, les botnets ont été surtout utilisés pour générer et propager des courriels indésirables. Ils peuvent être utilisés dans de nombreux buts malveillants, dont le vol de données personnelles et de mots de passe, l'attaque de réseaux publics et privés, l'exploitation de la puissance informatique et de l'accès à Internet des utilisateurs, et la mise en œuvre d'attaques de déni de service distribué (DDoS).1 En résumé, les botnets sont un problème complexe et en évolution continuelle qui constitue une menace à la confiance des utilisateurs en l'Internet. Diverses techniques sont utilisées pour infecter les ordinateurs pour en faire des bots, entre autres convaincre les utilisateurs de télécharger des logiciels malveillants, exploiter les vulnérabilités des navigateurs, persuader les utilisateurs d'enregistrer un logiciel malveillant (par ex. à la suite de l'ouverture d'une pièce jointe à un e-mail infectée). Le logiciel malveillant d'un botnet est conçu pour fonctionner en arrière plan, de sorte que les utilisateurs ne savent pas que leurs systèmes sont infectés. Bien que les botnets constituent une menace aux utilisateurs de l'Internet et soient difficiles à éliminer, on peut prendre des mesures pour réduire leur impact et les risques associés.

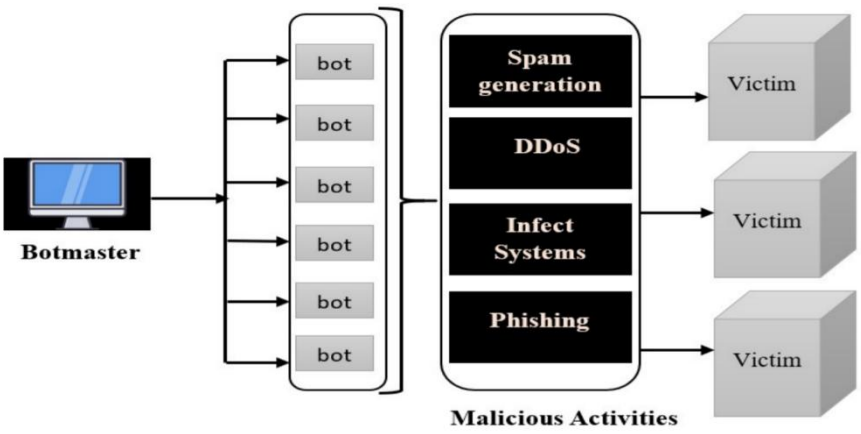
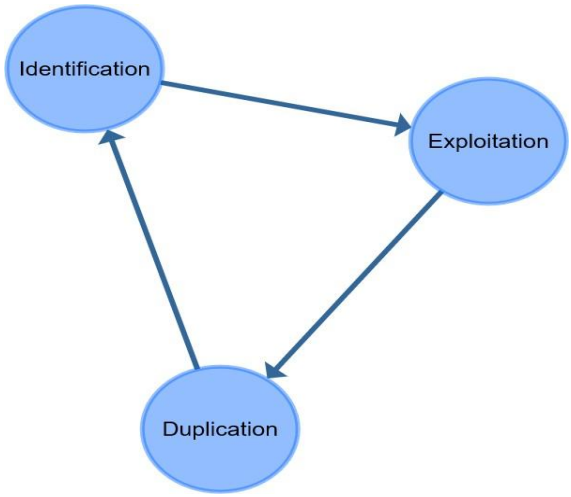


Figure II.2: Architecture de botnet

II.4 Méthodes de communication :

Le but principal d'un réseau de zombies est de grossir au maximum pour ensuite pouvoir lancer diverses attaques. Nous détaillerons ces dernières dans la prochaine section. Ici, nous allons décrire le fonctionnement d'un de ces réseaux, en nous basant sur le cycle de vie du programme Mirai. Tout d'abord, il faut identifier les objets vulnérables. Pour ce faire, les réseaux de zombies doivent constamment scanner l'ensemble des adresses IP disponibles afin de trouver le maximum de victimes. Ensuite, pour chaque victime potentielle, le réseau va essayer de l'exploiter. Ensuite, si l'attaque réussit, la victime va se faire infecter et elle participera au prochain cycle. Nous pouvons résumer ce cycle en trois étapes : identification des objets, exploitation des victimes et duplication du logiciel malveillant.



FigureII.3: Cycle defonctionnementgénérald'unréseaudezombiegrandissant

II.5 Type d'attaque :

Un botnet est plus dangereux que les menaces conventionnelles telles que les virus et les vers. Le projet Honeynet a présenté différents types d'attaques de botnet, y compris la fraude au clic, déni de service (DDoS), spam, cyberguerre, exploitation des ressources et vol information confidentielle [59]. D'autres études ont montré que les botnets peuvent être manipulésmener une grande variété d'activités illégitimes et plusieurs types de cybercriminalité.

II.5.1 Déni de service distribué :

Les botnets peuvent lancer des attaques par déni de service distribué (DDOS), dans lesquelles le trafic provenant d'un nombre énorme de sources inonde le système de la victime. Le nombre énormedes participants à un botnet confère au DDoS un pouvoir destructeur

important. Ceci permetbotmasters à utiliser le botnet pour détruire le système de contrôle de la victime en commandantmembres du bot pour envoyer un nombre massif de requêtes au système de la victime.

Jeux en ligneet les sites de jeux d'argent sont des exemples de cette attaque [60, 61, 62]. Globalement, le nombre deattaques est passé de 100 Gbps à 400 Gbps entre 2018 et 2019, et le nombre total d'attaques DDoS devrait doubler, passant de 7,9 millions en 2018 à 15,4 millionsd'ici 2023 [63].

II.5.2 Pourriel :

Les spams sont des messages électroniques indésirables qui contiennent souvent des liens ou des publicités malveillantsqui sont envoyées à un grand nombre d'utilisateurs. Pour un attaquant, un botnet est l'option la plus sûre pourutiliser comme plate-forme pour l'envoi de courriers indésirables. Le botnet « Grum » a envoyé environ40 milliards d'e-mails malveillants, utilisant les 600 000 bots de son réseau. Cetteattaque a commencé par envoyer des commandes botmaster aux bots jusqu'à ce qu'ils commencent à envoyer des spams à l'adressede la victime [64].

II.5.3 Voler des informations :

Un botmaster peut ordonner aux bots d'obtenir des données secrètes d'hôtes compromis en utilisantméthodes telles que la capture d'écran, la lecture des fichiers journaux et le keylogging. SDBot est un exemple debotnet qui utilise un logiciel avancé d'enregistrement de clés pour collecter des informations personnelles, quipeuvent ensuite être vendus à d'autres pour effectuer des actions illégitimes. Les méthodes d'enregistrement de frappe sont les principalesoutils utilisés par Zeus Bots pour voler des comptes bancaires privés et des informations de carte de crédit. Cettepermet au botmaster d'extraire les noms d'utilisateur et les mots de passe d'un compte de réseau social,le site Web de la banque et les e-mails. De plus, le bot peut récupérer des informations d'utilisateur privéesà partir de l'interface de programme d'application (API) Windows avant qu'il ne soit crypté par le Webnavigateur.

II.5.4 Exploiter les ressources :

Les hôtes compromis sont recrutés pour exécuter des actions illégales. Par exemple, les robots étaientutilisé sur Twitter et Facebook pour émettre de faux votes et augmenter le nombre de followers.De plus, un bot peut utiliser la machine de la victime pour accéder à un site Web lors d'une visite régulière àaugmenter le nombre d'utilisateurs du site Web sans leur autorisation.

Les botnets sont utilisés pour lancer diverses formes/types d'attaques DDoS, telles que les attaques de couche application (attaques basées sur HTTP), les inondations SYN et les

attaques d'amplification DNS. Le bot a été chargé de submerger le système cible avec de grandes quantités de trafic rapide (par exemple, des requêtes HTTP, des paquets SYN et des requêtes DNS).

i. Inondation DNS envoie des requêtes DNS falsifiées à un débit de paquets élevé et à partir d'un large éventail d'adresses IP sources vers le réseau cible. Étant donné que les demandes semblent valides, les serveurs DNS de la victime répondent à toutes les demandes falsifiées et leur capacité peut être dépassée par le nombre de demandes. Cette attaque consomme de grandes quantités de bande passante et d'autres ressources du réseau. Finalement, il épuise l'infrastructure DNS jusqu'à ce qu'il tombe en panne, prenant l'accès Internet de la victime et l'hébergement hors ligne sites avec. [65]

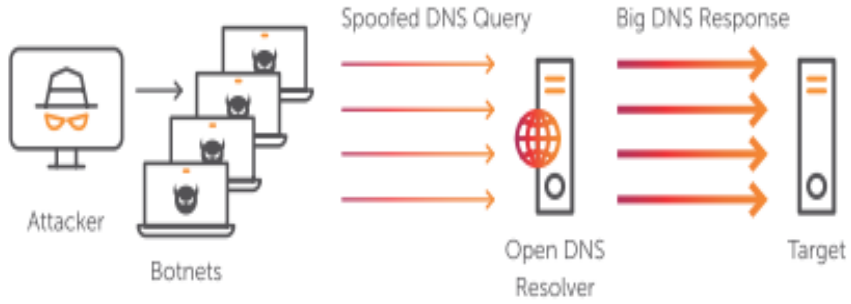


Figure II.4: Inondation DNS. [19]

ii. Inondation SYN , souvent généré par des bot nets, est conçu pour consommer les ressources de la victime serveur, tel qu'un pare-feu ou un autre périmètre éléments de défense, dans une tentative de submerger ses limites de capacité et l'abaisser. La cible reçoit des paquets SYN à des débits très élevés qui remplir rapidement sa table d'état de connexion, résultant dans les déconnexions, la chute du trafic légitime paquets, ou pire encore – redémarrage de l'élément. [.65].

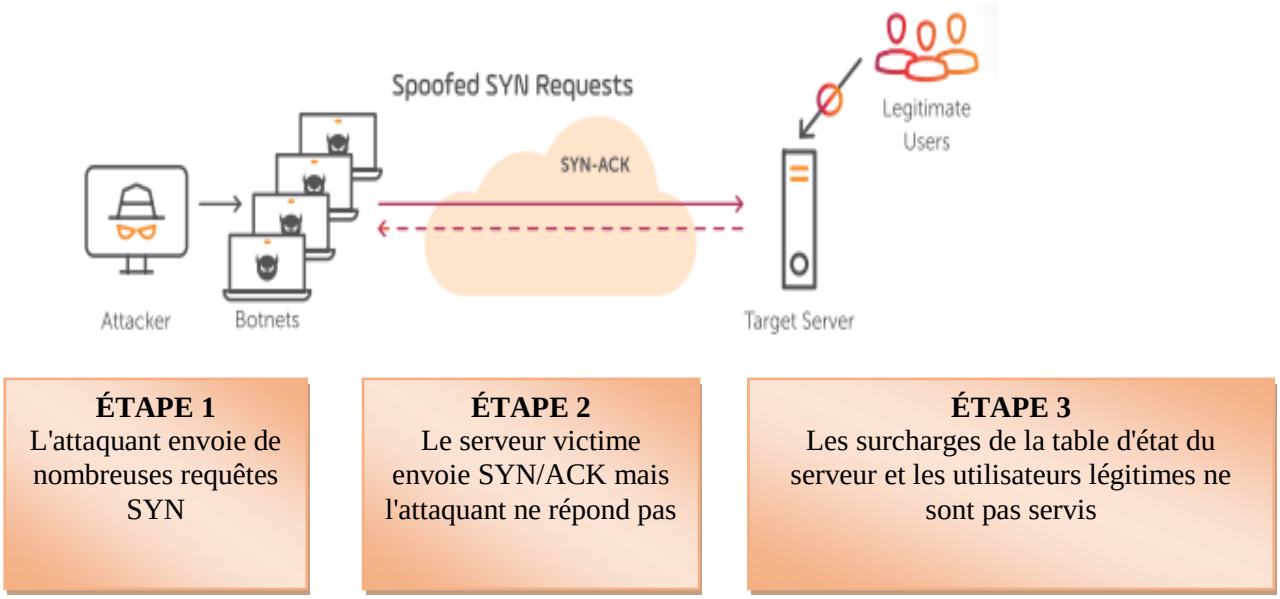


Figure II.5: Inondation SYN. [65]

iii. Inondation http : HTTP (et sa forme cryptée HTTPS) est un protocole de transport pour Internet basé sur un navigateur requêtes, couramment utilisées pour charger des pages Web ou pour envoyer du contenu de formulaire sur Internet. Dans une attaque par inondation HTTP/S exploitée par l'attaquant HTTP GET ou post apparemment légitime demandes d'attaque d'un service Web ou d'une application.

Ces attaques utilisent souvent de nombreux bots nets tels qu'appareils IoT infectés. Les appareils sont coordonnés pour envoyer plusieurs GET demandes de fichiers image ou d'autres actifs du serveur Web cible. Le flot de requêtes http épuise les ressources du serveur jusqu'à ce qu'un déni de service se produise pour les demandes provenant d'utilisateurs légitimes.

Un flot HTTP peut également être lancé en envoyant plusieurs requêtes post qui déclencheront un traitement intensif traitement sur le serveur et saturera le serveur ressources encore plus rapidement. [65]

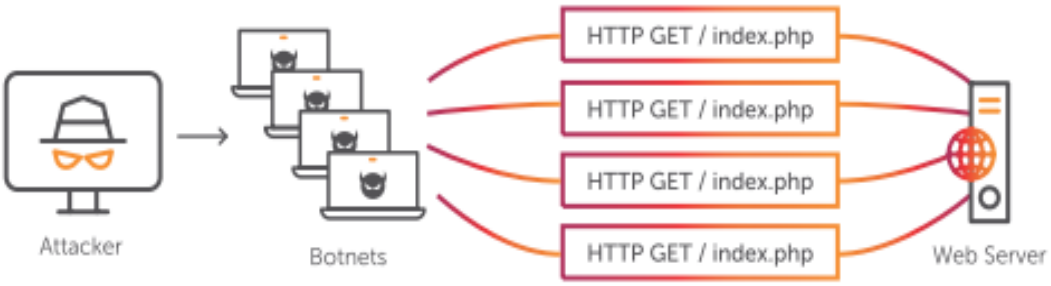


Figure II.6 : Inondation http. [65]

II.6 Fraude au clic :

Le bot master l'utilise pour générer de faux clics pour des publicités en ligne qui imitent un mode de requête légitime (généralement en utilisant des points de terminaison d'en-tête de requête HTTP), obligeant les annonceurs à payer beaucoup d'argent. La publicité en ligne devient très populaire, et le modèle de tarification de ce type de publicité est généralement basé sur une approche de paiement au clic, ce qui signifie que les revenus des plateformes publicitaires (telles que Facebook, Google) dépendent du nombre de clics sur la plateforme publicitaire. Malheureusement, certains hackers profitent de ce modèle et utilisent des bot nets pour effectuer des clics frauduleux.[67]

II.7 Architecture de botnet :

Les botnets sont classés en trois structures différentes selon le canal C&C : architecture centralisée, architecture décentralisée et architecture hybride [68, 69– 70]. Ac-

Selon les protocoles de communication utilisés par les botnets, les botnets peuvent être classés enplusieursprotocoles.discutent dans le tableau 1

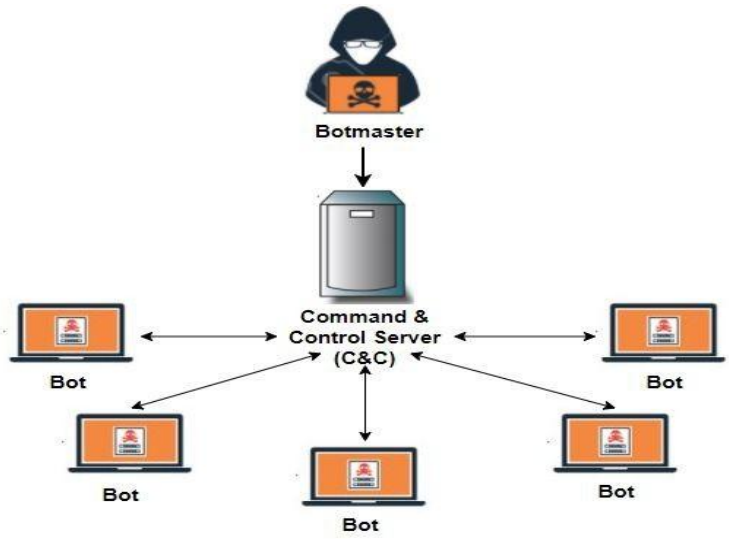
Tableau II .1 Botnetprotocols.

Protocols	Def	Advantage	Examples
IRC	IRCIRC is a protocol of real-time internet textmessagingchat;Mainlyusedincentral-izedarchitecture.	Low-latencycommunication. 1. Simplecommands. 2. Private(one-to-one)communication. 3. Capableofgroup(manyto-many)commu-nication. 4. simpletosetup. 5. Flexibilityincommunication. 6. Anonymousreal-timecommunication	Agobot, SDBot,Spybot,andGTBot
HTTP	HTTP protocols attempt to blend botnettraffic into regular HTTP traffic. Mainlyusedincentralizedarchitecture.	Difficulttodetectandeasilybypassesfire-walls.	Bobax, ClickBot,RustockandBlackenergy.
P2P	P2Pisacommunicationprotocolwhichismainlyusedindecentralizedarchitecture	hardtodetect,veryhighresilience.	Slapper,Sinit, Phatbot,Nu-gache,Stor

II.7.1 Architecture centralisée:

Dans une architecture de botnet centralisée, le botmaster contrôle tous les bots à partir d'unhub connu sous le nom de serveur de commande et de contrôle. Dans cette structure, un seul point (le C&Cserver) est utilisé pour échanger des instructions entre le botmaster et les bots. Le principall'avantage de cette architecture est qu'elle fournit une coordination fiable des bots pour leurbotmaster. De plus, cela facilite la surveillance de l'état pour le botmaster et accélèretemps de réaction. En revanche, une fois le serveur C&C identifié, il est très facile pour un défenseurpour éliminer ce type de botnet. Les deux protocoles les plus souvent

utilisés dans un archivage centralisésont le chat par relais Internet (IRC) et le protocole de transfert hypertexte (HTTP). Un centralisél'architecture peut subir un point de défaillance unique, en raison d'une attaque par déni de service, etlebotmaster n'est plus en mesure de communiquer avec les bots lorsqu'un serveur IRC ou HTTPest retiré [68, 69– 71]. La figure 1 montre un modèle centralisé



FigureII.7 architecture centralisé.

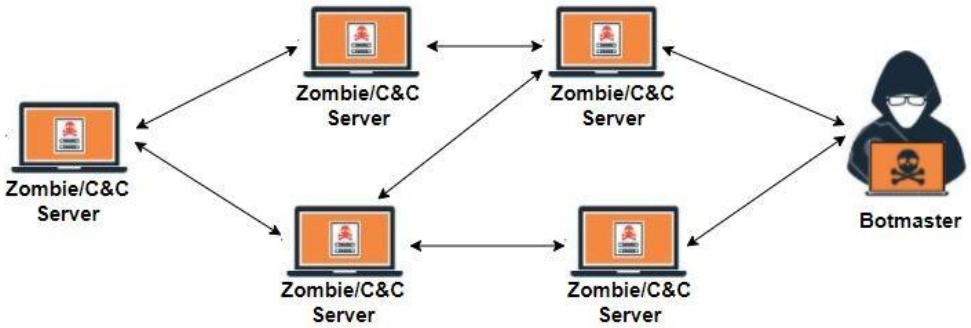
Botnet basé sur IRC : Le chat relais Internet (IRC) est un protocole d'Internet en temps réelmessagerie texte ou conférence synchrone. Ce protocole est le plus populaire canal C&C du botnet. Les raisons de cette popularité sont les suivantes : temps de communication avec les bots, il prend en charge un grand nombre de clients, il fonctionneavec différentes topologies de réseau, c'est un open-source et il a une conception extensible.Spybot, Agobot, SDBot et GT Bot sont les plus célèbres botnets basés sur IRC.

Botnet basé sur HTTP : Le protocole de transfert hypertexte (HTTP) est un autre protocolutilisé par les serveurs C&C. La communication HTTP est largement utilisée dans de nombreuxapplications. Ces robots C&C basés sur le Web tentent de se fondre dans le trafic HTTP régulier, les rendant difficiles à détecter. Ils peuvent facilement échapper aux systèmes de détection d'intrusion (IDS) et contourner les pare-feu avec des techniques de filtrage basées sur les ports. Bobax, ClickBot, Rustock, et le plus populaire, Blackenergy, sont des bots bien connus qui utilisent le protocole HTTP.

II.7.2 Architecture décentralisée :

Une forme décentralisée permet aux bots d'agir de manière autonome. Dans cette structure, le

système de communication n'est pas basé sur des serveurs pour détruire et découvrir un certain nombrede bots, et il n'y a pas de point de communication centralisé. Dans ce genre de botnet, chaquebot peut établir des connexions avec d'autres bots, et les bots se comportent à la fois comme des serveurs et des clients.Peer to Peer (P2P) est le protocole le plus utilisé dans une architecture centralisée, la figure 8 montre unmodèle décentralisé [68, 69– 70].



FigureII.8 architecture décentralisé.

Botnet basé sur Peer to Peer (P2P) : Ce protocole de communication est principalement utilisé dansune architecture décentralisée. L'objectif des botnets P2P est de supprimer le point de défaillance, qui est la principale vulnérabilité d'une structure centralisée. Par conséquent, la détection de cetype de botnet est très difficile. Pour envoyer des commandes à tous les bots sur l'ensemble du réseau, le botmaster doit se connecter à un seul des bots (pairs) [68, 72, 71]

II.7.3 Architecture hybride :

La combinaison d'architectures centralisées et décentralisées est une architecture hybride.Unbotnet hybride est divisé en deux types de bots : l'un est un serviteur et l'autre est un clientrobot. Le bot serviteur reçoit les commandes du botmaster et les transmet au client robot. Il est plus difficile de détecter et de contrôler les botnets dans un réseau à architecture hybridequ'avec les architectures centralisées ou décentralisées, pourtant la conception du botnet n'est pas très complexe [68, 73, 70].

II. 8 Le cycle de vie d'un botnetiot :

L'apprentissage du cycle de vie des botnets est un facteur important dans l'analyse réussie des systèmes de détection de botnets. Comprendre chaque phase de ce cycle peut aider à améliorer et à développer un système de détection de botnet efficace. L'hôte doit passer par cinq phases pour devenir un bot actif et faire partie d'un botnet comme décrit dans les références, la figure 9 montre le cycle de vie d'un botnet [74 , 75 , 76 – 77]

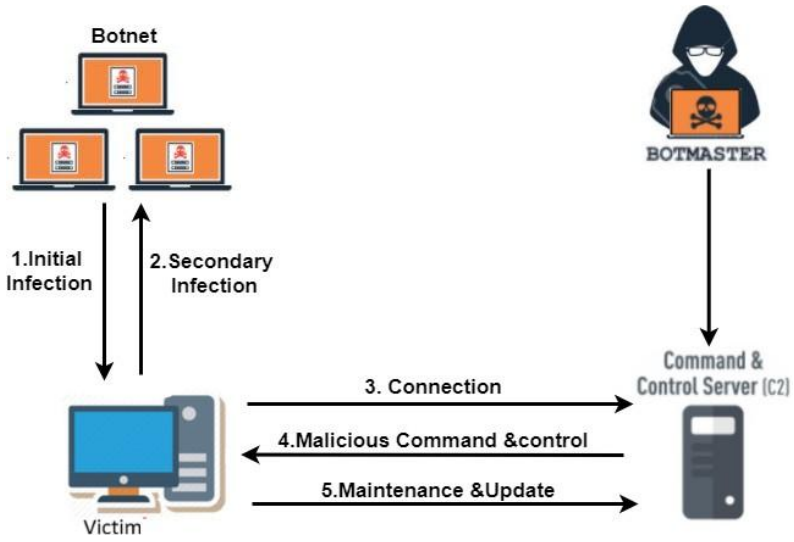


Figure II.9 Cycle de vie du botnet.

La première étape est appelée la première injectionde phase, où un hôte est compromise et devient un bot potentiel. Un hôte peut être infecté par différentes techniques telles que l'exécution de logiciels malveillants, le téléchargement de logiciels malveillants à partir d'un site Web non autorisé, le clic sur une pièce jointe d' un e-mail ou l'utilisation d'un disque amovible infecté tel qu'une clé USB ou une clé USB. Dans la deuxième phase, également appelée phase d'injection secondaire, l'hôte compromis exécute un programme dans une base de données réseau spécifiée qui recherche les fichiers binaires de logiciels malveillants. Une fois le programme de bot installé, l'hôte infecté devient un véritable bot (ou zombie) et exécute le code malveillant. Les binaires Bot sont généralement téléchargés à l'aide des protocoles FTP ou HTTP. La troisième phase est également appelé phase de connexion ou de ralliement, où un canal C&C est installé par le programme de bot et commence ensuite à communiquer entre le bot et le serveur C&C. Une fois connecté, le bot peut recevoir et répondre aux commandes du botmaster, et la machine infectée devient une partie de l'armée du botnet de l'attaquant. Cette phase se produit plusieurs fois dans le cycle de vie du bot [74, 75, 76 – 77].La quatrième phase est aussi appelée la phase malicieuse, où la commande réelle et contrôle

les opérations de l' botnet sont commencé. Le bot tente d'exécuter une série de malveillants o pérations en fonction de la de Botmaster instructions. Il s'agit notamment du spam, de la fraude d'identité, des fuites d'informations, du déni de service distribué (DDoS) et de la fraude au clic. La phase finale est également appelée phase de maintenance et de mise à niveau, où des bots mis à jour sont utilisés pour garder les autres bots sous le contrôle du

botmaster. Les bots ont besoin de cette phase pour plusieurs raisons. Ils peuvent avoir besoin d'échapper à des techniques de détection en mettant à jour le binaire bot, ou ils peuvent vouloir d'ajouter de nouvelles fonctionnalités à leur bot armée [74 , 75 , 76- 77]

II.9 Classement des réseaux de zombies en fonction du protocole de réseau utilisé :

Afin de guider le bot, le maître du bot net doit au moins établir une connexion réseau entre le bot et l'ordinateur qui envoie les instructions. Toute coopération en réseau est basée sur le protocole réseau, qui définit les règles de communication informatique dans le réseau. C'est pourquoi les bots nets sont classés selon le protocole de communication utilisé.
Selon le protocole réseau utilisé, les bots nets peuvent être classés dans les catégories suivantes :

II.9.1 Orientation IRC :

C'est l'un des premiers types de bot nets à gérer des robots via IRC (Internet Relay Chat). Chaque ordinateur infecté se connectera au serveur IRC indiqué dans le corps du robot, entrera dans un canal défini et attendra les instructionsdu maître. [78]

II.9.2 Orientés vers les messageries instantanées :

Ce type de bot net n'est pas courant. La seule chose qui le distingue de ses homologues orientés IRC est l'utilisation de canaux de messagerie instantanée (AOL, MSN, ICQ, etc.) pour transmettre des données.

Ces bots nets sont impopulaires car il est difficile de créer un nouveau compte de messagerie instantanée pour chaque bot. Le robot doit être sur le réseau et toujours en ligne. Dans la mesure où la majorité des administrations de messagerie instantanée ne permettent pas l'association depuis différents ordinateurs à l'aide du même compte, chaque bot doit posséder propre numéro d'assistance de messagerie instantanée. Et les fournisseurs d'administrations de messagerie instantanée empêchent les enregistrements automatiques de comptes. Standard conséquent, le propriétaire d'un réseau de zombies orienté vers les messageries instantanées est limité au niveau du nombre de comptes enregistrés et, standard conséquent, au niveau du nombre de bots qui peuvent être en ligne simultanément. Bien entendu, les bots peuvent utiliser le même compte et se connecter une fois à un second déterminé, envoyer les données au numéro du propriétaire et attendre la réponse pendant un laps de temps défini mais cette méthode n'est pas sans problèmes : un réseau de ce genre réagit très lentement aux guidelines. [78].

II.9.3 Orientés vers Internet :

Les réseaux de zombies qui basent leur administration via Web sont relativement récents et se développent rapidement. Le bot se connecte à un serveur Web spécifique, reçoit des instructions et transmet des données en réponse. Ces botnets sont populaires en raison de leur facilité de développement et de gestion via une interface Web, du grand nombre de serveurs Web sur Internet. [78]

II.9.4 Autre :

En plus des catégories énumérées ci-dessus, il existe d'autres types de botnets qui utilisent leur propre protocole heuristique TCP/IP pour communiquer : ils n'utilisent que les protocoles communs TCP, ICMP et UDP [78]

Chapitre III: Machine Learning

III.1 Introduction:

Le machine learning, que l'on traduit en français par « apprentissage automatique », ou plus généralement l'intelligence artificielle dont le machine learning est un sous domaine, évoque de la science-fiction pour la plupart des gens. Pourtant le machine learning n'est pas un rêve futuriste mais fait déjà partie de la vie quotidienne. A vrai dire, il s'est déjà imposé depuis des décennies comme dans les filtres anti-spam vers les années quatre-vingts dix ou encore dans les jeux vidéo depuis le début des années 2000. Aujourd'hui le machine learning est partout, il existe que ce soit à travers la voiture autonome de Google, la reconnaissance de parole de Siri, la detection de visage de Facebook ou encore le magasin autonome d'Amazon. Le machine learning est en train de révolutionner le monde

III.2 Définition:

L'apprentissage automatique (en anglais : *machine learning*, litt. « apprentissage machine »), apprentissage artificiel ou apprentissage statistique est un champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d'« apprendre » à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune. Plus largement, il concerne la conception, l'analyse, l'optimisation, le développement et l'implémentation de telles méthodes.[79]

L'apprentissage automatique comporte généralement deux phases. La première consiste à estimer un modèle à partir de données, appelées observations, qui sont disponibles et en nombre fini, lors de la phase de conception du système. L'estimation du modèle consiste à résoudre une tâche pratique, telle que traduire un discours, estimer une densité de probabilité, reconnaître la présence d'un chat dans une photographie ou participer à la conduite d'un véhicule autonome. Cette phase dite « d'apprentissage » ou « d'entraînement » est généralement réalisée préalablement à l'utilisation pratique du modèle. La seconde phase correspond à la mise en production : le modèle étant déterminé, de nouvelles données peuvent alors être soumises afin d'obtenir le résultat correspondant à la tâche souhaitée. En pratique, certains systèmes peuvent poursuivre leur apprentissage une fois en production, pour peu qu'ils aient un moyen d'obtenir un retour sur la qualité des résultats produits.[80]

III.3 Les différents types d'apprentissage

Il existe plusieurs types de système d'apprentissage et cela varie en fonction du type de problème que l'on se pose. Il est alors utile de les classer en différentes catégories. Les systèmes de machine learning peuvent-être classés en fonction de l'importance et de la nature de la supervision qu'ils requièrent durant la phase d'entraînement. On distingue alors quatre grandes catégories: l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage semi-supervisé et l'apprentissage avec renforcement.

III.3.1 Apprentissage Supervisé :

L'apprentissage supervisé consiste en la conception d'un modèle reliant des données d'apprentissage à un ensemble de valeurs de sortie. C'est-à-dire que les données d'entraînement qu'on fournit à l'algorithme comportent les solutions désirées, appelées étiquettes (en anglais, labels). Cette méthode permet donc à l'algorithme d'apprendre en comparant sa sortie réelle avec les sorties enseignées, afin de trouver les erreurs et modifier le modèle en conséquent. L'apprentissage supervisé confère au modèle la possibilité de prédire des valeurs d'étiquette sur des données non étiquetées supplémentaires. Soit D un ensemble de données, décrit par un ensemble de caractéristiques X, un algorithme d'apprentissage supervisé va trouver une fonction de mapping entre les variables prédictives en entrée X et la variable à prédire Y. La fonction de mapping décrivant la relation entre X et Y s'appelle un modèle de prédiction. Les caractéristiques X peuvent être des valeurs numériques, alphanumériques ou des images. Un exemple d'utilisation de l'apprentissage supervisé est le filtre anti-spam, l'apprentissage s'effectue à l'aide de nombreux exemples de-mails qu'on a étiqueté spam ou normal. A partir de cela, le filtre doit alors être capable de classer de nouveaux e-mails. [81] [82] [83]

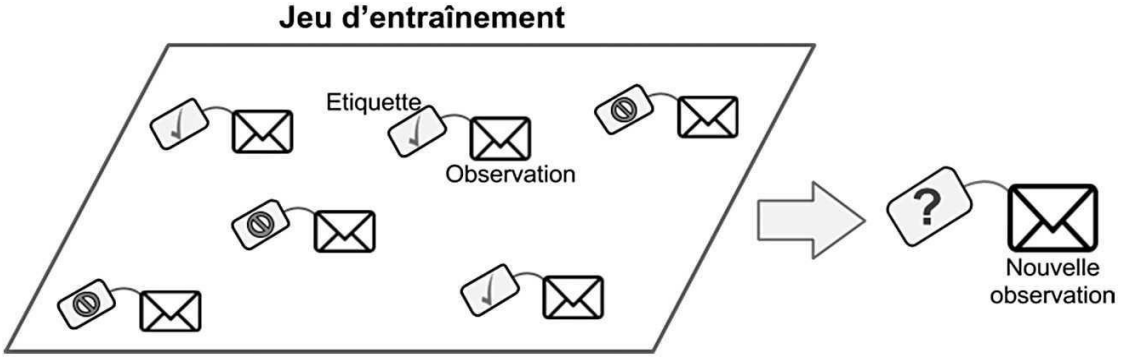


Figure 3.01 **:** Un jeu d'entraînement étiqueté pour un apprentissage supervisé

III.3.2 Apprentissage non Supervisé :

L'apprentissage non supervisé consiste en la conception d'unmodèle structurant l'information, c'est-à-dire les données d'apprentissage ne sont pas étiquetées. Cette méthode permet donc à l'algorithme de trouver tout seul des points communs parmi les données d'entrée, le système apprend alors sans professeur. Comme l'étiquetage de données requiert beaucoup de temps, les méthodes d'apprentissage utilisant l'apprentissage non supervisé sont particulièrement utiles.

L'apprentissagenonsupervisépeut êtreutilisépour laréductiondedimension oul'extractionde variable. Cette tâche consiste à simplifier les données sans perdre trop d'informations, ilpourra ensuite être fourni à un autre algorithme d'apprentissage automatique (tel qu'un algorithmed'apprentissagesupervisé).Lekilométraged'unevoiture, parexemple, peutêtrefortementcorréléàson âge, de sorte que l'algorithme de réduction de dimension les combinera en une seule variable représentant la vétusté de lavoiture.

A première vue, on pourrait penser que l'apprentissage non supervisé a peu d'utilité dans lesapplications de la vraie vie, mais les applications de cette technique sont nombreuses. Les sites comme Amazon, Netflix ou encore Youtube utilisent les algorithmes de partitionnement ou Clustering en anglais pour faire des recommandations de produits ou de films. Il peut être également utilisé pour explorer de larges ensembles de données et de découvrir d'intéressantes relationsentrelesvariables:pourunsupermarchéparexemple, exécuterunerègled'associatio urles journaux de vente permettrait peut-être de découvrir que les personnes achetant de la sauce barbecue et des chips ont aussi tendance à acheter des grillades. Cela permettrait de réorganiser les rayons afin de présenter ces articles à proximité les uns des autres. [84] [85][86]

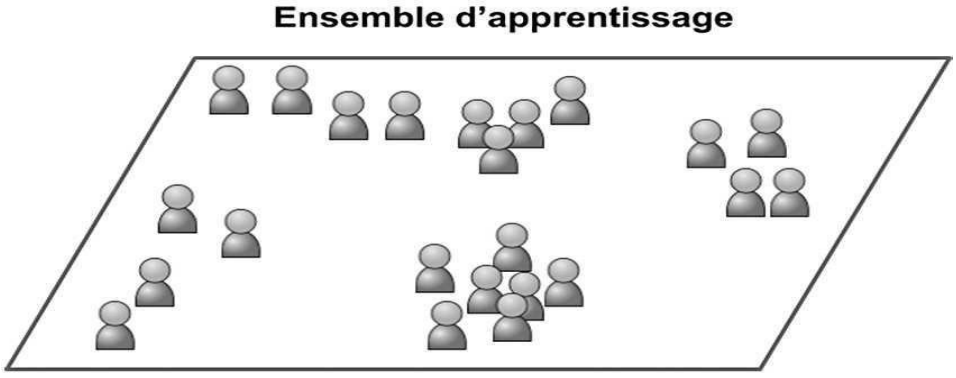


Figure 3.02 **:** Un jeu d''entraînement non étiqueté pour un apprentissage non supervisé

III.3.3 Apprentissage par renforcement :

L'apprentissage se fait sans supervision, par interaction avec l'environnement (principe d'essai/erreur) et, en observant le résultat des actions prises.Chaque action de la séquence estassociée à une récompense. Le but est de déterminer la stratégie comportementale optimale afin de maximiser la récompense totale. Pour cela, un simple retour des résultats est nécessaire pour apprendrecommentlamachinedoitagir.Ceciestappelélesignalderenforcement.Ilpeutêtre très avantageux pour la prévision financière à haute fréquence où l'environnement est dynamique et en conséquence, il est difficile de trouver ou d'automatiser manuellement des stratégies efficaces.

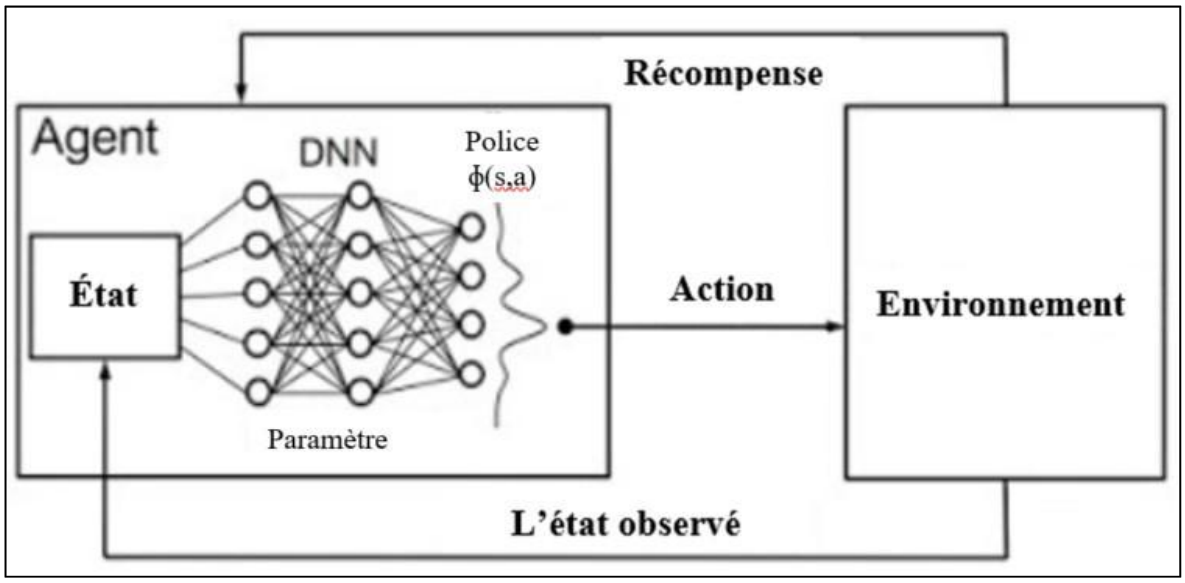


Figure 3.3schéma descriptive de l'apprentissage parrenforceme

III.4 Types des algorithmes d'Apprentissage Automatique

Les algorithmes d'apprentissage peuvent se catégoriser selon le mode d'apprentissage qu'ils emploient :

III.4.1 LESk-PLUS PROCHES VOISINS (k-NN)

L'algorithme des k-plus proches voisins (k-NN) se base sur les données en entier. Eneffet,pour uneobservation, quinefait pas partiedesdonnées, qu'on souhaiteprédire,l'algorithme va chercher les k instances les plus proches de notre observation et choisir pourchaque observationla classemajoritaireparmiseskplusprochesvoisins.La méthode k-NN est une technique d'apprentissage supervisé, et est

considéréecommel'unedesplussimplesdansledomainedelaclassification.Ellepermetdeclassifier unenouvelle observation (vecteur de caractéristiques extraites) en calculant la distance avec lesdonnéesd'entraînement,etdeprendreleskplusprochesvoisins(entermesdedistance).Puis,observé la classe qui est majoritairement représentée parmi les k-plus proches voisins etd'assignercetteclasse àlanouvelleobservation.Bienqueletempsd'apprentissagedel'algorithmek-NNsoitcourt,letempsderequêteréel (et l'espace de stockage) peut être plus long que celui des autres modèles. Cela estparticulièrementvrailorsquelenombredepointsdedonnéesaugmente,car toutes les données d'entraînementdoivent être conservées,mais passeulementl'algorithme.
Le plus grand inconvénient de cette méthode est qu'elle peut être erronée par desattributs non pertinents qui masquent des attributs importants. Il existe des moyens de corrigerce problème, par exemple en appliquant des pondérations aux données. Comme nous l'avonsdétailléci-dessus, l'algorithmek-NNcalculeladistanceentrelespointsdedonnées.Pourcela, nous utilisons la formule dela distanceeuclidienne:

$$d(p,q)=d(q,p)=\sqrt{(q_1-p_1)^2+(q_2-p_2)^2+\cdots+(q_n-_n)^2}$$

$$d(p,q)=d(q,p)=\sqrt{\sum_{i=1}(q_i-p_i)}$$

(3.15)

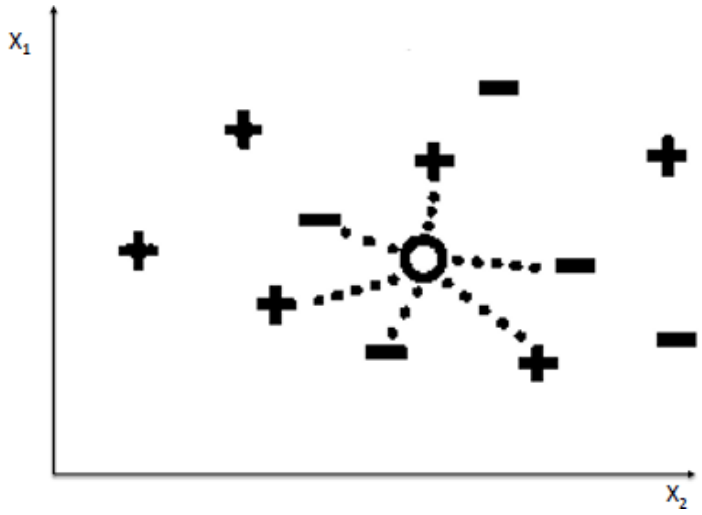


Figure3.4:Principedel'algorithmek-NN

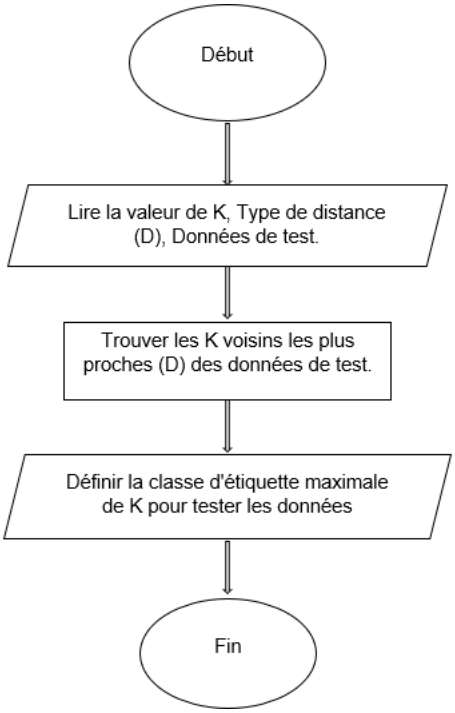


Figure 3.5:Organigrammedel'algorithmeKNN.

III.4.2 Les machines à support de vecteurs (SVM)

Les machines à vecteurs de support (SVM) ont été développées à l'origine par Vapnik et ses collègues en 1992 sur la base de la théorie de l'apprentissage statistique de Vapnik & Chervonenkis dans les années 1960. Le SVM a été appliqué avec succès dans de nombreuses applications, notamment la reconnaissance manuscrite, la prédiction de séries chronologiques, la reconnaissance vocale, le problème de séquence protéique, le diagnostic du cancer du sein et bien d'autres [87]

Les machines à vecteur de support (SVM) sont utilisées lorsque les données ont exactement deux classes. L'algorithme SVM classe les données en trouvant le meilleur hyperplan qui sépare tous les points de données d'une classe de ceux de l'autre classe (le meilleur hyperplan pour un SVM est celui avec la plus grande marge entre les deux classes). L'algorithme SVM peut aussi être utilisé avec plus de deux classes, auquel cas le modèle créera un ensemble de sous-ensembles de classification binaire. Il y a quelques avantages importants à utiliser l'algorithme SVM. Tout d'abord, il est extrêmement précis et n'a pas tendance à sur adapter les données. Deuxièmement, les machines à vecteurs de support linéaire sont relativement faciles à interpréter. Parce que les modèles SVM sont très rapides, une fois que votre modèle a été formé, vous pouvez supprimer les données de formation si vous avez une capacité de

mémoire disponible limitée. Il a également tendance à très bien gérer les classifications complexes et non linéaires en utilisant une technique appelée « astuce du noyau ». Cependant, les algorithmes SVM doivent être formés et réglés à l'avance, vous devez donc investir du temps dans le modèle avant de pouvoir commencer à l'utiliser. De plus, sa vitesse est fortement affectée si vous utilisez le modèle avec plus de deux classes.

L'algorithme SVM est un classificateur dit linéaire, ça veut dire que, dans le cas parfait, les données doivent être linéairement séparables. Il permet de trouver le meilleur séparateur (ligne, plan ou hyperplan) qui sépare le mieux les deux classes

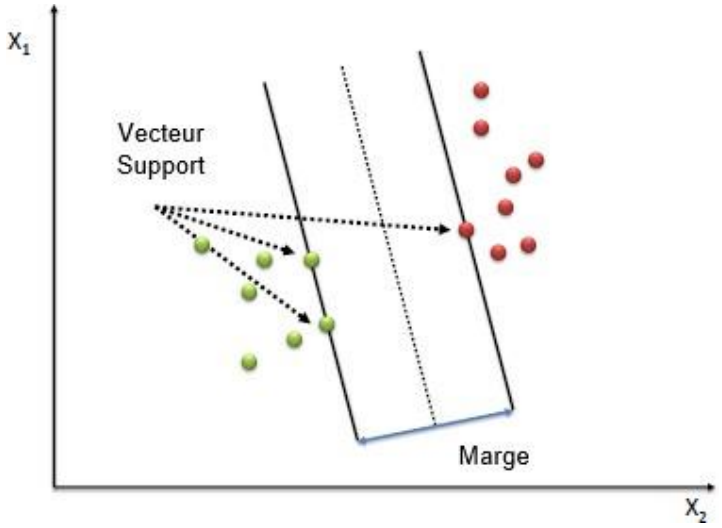


Figure 3.6:Principedel'algorithmeSVM.

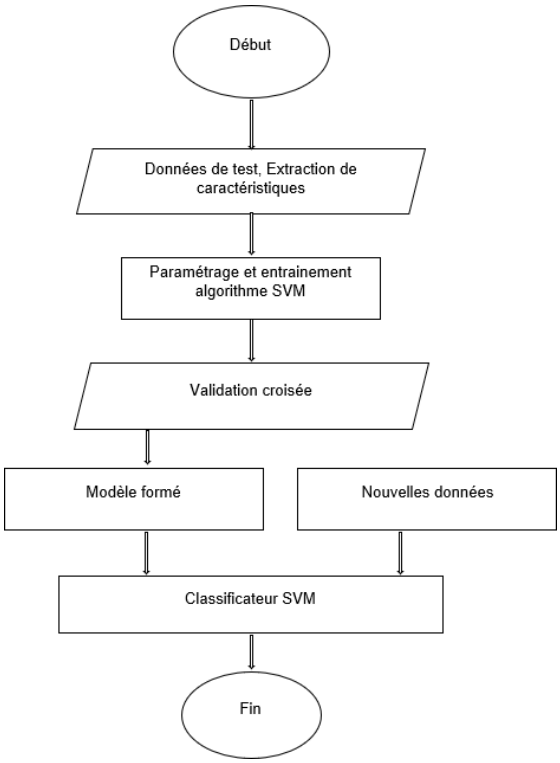


Figure3.7:Organigrammedel'algorithmeSVM

III.4.3 Arbres de décision

L'arbre de décision est l'une des méthodes souvent utilisées pour la classification et la prédiction. Un arbre est formé, où chaque nœud de branche représente un choix entre un certain nombre d'alternatives et chaque nœud de feuille représente une décision. Les règles basées sur les données sont représentées par la structure séquentielle des arbres de décision et elles partitionnent les données de manière itérative.

Elle se compose d'un nœud racine, le nœud supérieur de l'arbre comprenant toutes les données, d'un nœud de division, un nœud qui divise les données entre les alternatives et d'un nœud terminal (ou feuille), un nœud où le résultat ou la décision finale est disponible.

Chacun des nœuds internes d'un arbre de décision divise l'espace d'instance en deux sous-espaces ou plus selon une certaine fonction discrète des valeurs des attributs d'entrée. Si les attributs contiennent une valeur numérique, chaque feuille est affectée à une classe représentant la valeur cible la plus appropriée. Les instances sont classées en les faisant naviguer de la racine de l'arbre vers une feuille, en fonction du résultat des tests effectués le long du chemin.[88]

ID3, CART et C4.5 sont les trois algorithmes d'apprentissage par arbre de décision les plus utilisés.

- L'algorithme ID3 est un algorithme de construction d'arbres de décision.Il détermine la

classification des objets en testant les valeurs des propriétés.

- CART est un autre algorithme, qui signifie arbres de classification et de régression. Pour construire un arbre de décision, il construit des arbres binaires.
- L'algorithme C4.5 génère un arbre de décision pour les données fournies en divisant récursivement ces données et .il considère tous les tests possibles qui peuvent diviser les données et sélectionne un test qui donne le meilleur gain d'information.

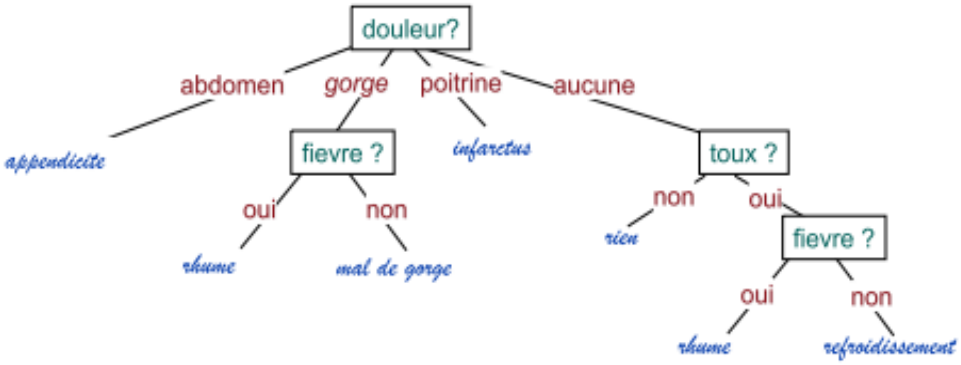


Figure 3.8:construction d'un arbre de décision article [41]

III.4.3.1 Avantages :

- Il est Facile à comprendre [89]
- Il est Facile à interpréter [.89]
- c'est un algorithme assez simple qui n'est pas très coûteux en temps de calcul.[89]

III.4.3.2 Inconvénients :

- Ce type d'algorithmes est très sensible aux points aberrants et au bruit [90]

III.4.3.3 Foret aléatoire :

La forêt aléatoire ou forêt de décision aléatoire est un ensemble de méthodes d'apprentissage utilisées pour la classification, la régression et d'autres tâches. C'est également l'algorithme le plus exible et le plus facile à utiliser. Une forêt est composée d'arbres. On dit que plus il y a d'arbres, plus la forêt est robuste

Son principe de fonctionnement est de construire un grand nombre d'arbres de décision

Sur des échantillons de données sélectionnés au hasard, obtiennent des prédictions à partir de chaque arbre et sélectionnent la meilleure solution par vote. Il fournit également un assez bon indicateur de l'importance des fonctionnalités [91]

La forêt aléatoire a de nombreuses applications, telles que les moteurs de recommandation, classification des images et sélection des caractéristiques. Il peut être utilisé pour classer les

demandeurs de prêt identifient les activités frauduleuses et prédisent la maladie.

Nous pouvons comprendre le fonctionnement de l'algorithme Random Forest à l'aide des étapes suivantes :

- Étape 1 : Commencez par sélectionner des échantillons aléatoires à partir d'un ensemble de données.
- Étape 2 : Ensuite, cet algorithme construira un arbre de décision pour chaque échantillon. Ensuite, il obtiendra le résultat de la prédiction de chaque arbre de décision.
- Étape 3 : Dans cette étape, le vote sera effectué pour chaque résultat prévu.
- Étape 4 : Enfin, sélectionnez le résultat de prédiction le plus voté comme résultat de prédiction final [92]

Le schéma suivant illustre son fonctionnement

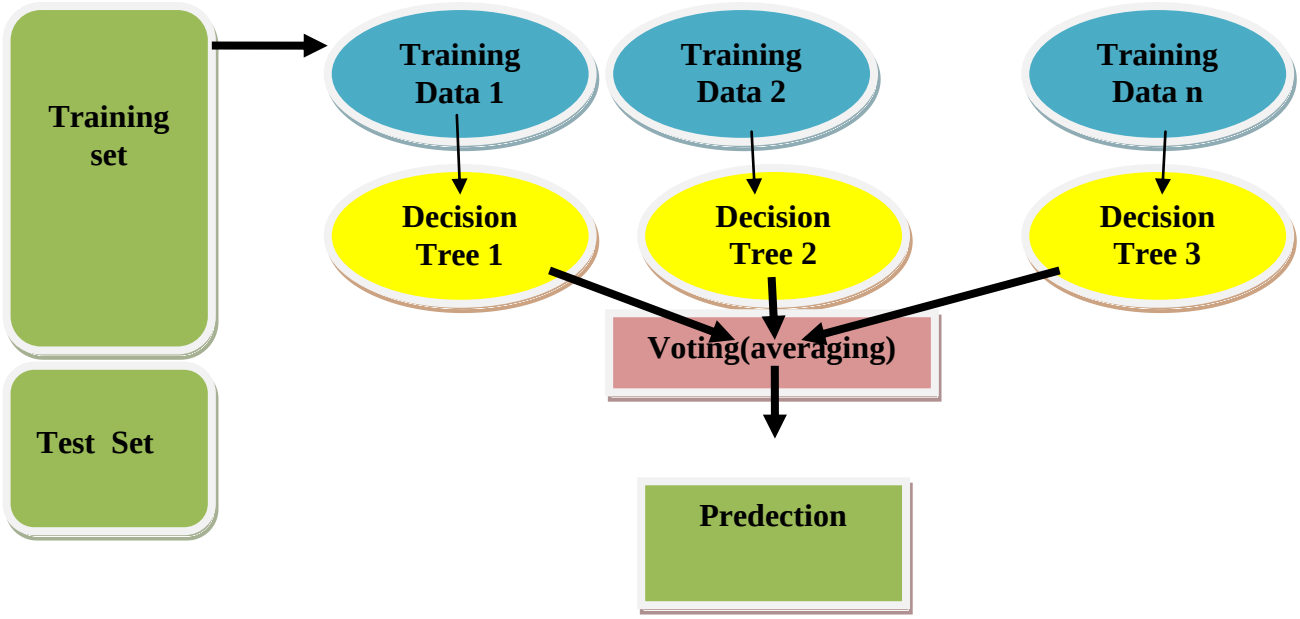


Figure3.9:construction d'un foret aléatoire [14]

III.4.5 La régression logistique

Il s'agit d'une méthode statistique d'analyse d'un ensemble de données dans lequel se trouvent un ou plusieurs variables qui déterminent un résultat. Le résultat est mesuré avec une variable dichotomique (en où il n'y a que deux résultats possibles). Le but de la régression logistique est de trouver le meilleur modèle d'ajustement pour décrire la relation entre la caractéristique dichotomique d'intérêt (variable dépendante = variable de réponse ou de résultat) et un ensemble de variables indépendantes (prédicteur ou variables explicatives).[93]

Généralement, la régression logistique veut dire une régression logistique binaire ayant des variables binaires, mais il peut y avoir deux autres catégories de variables cibles qui peuvent être prédites par elle.La régression logistique peut être divisée en types suivants : [94]

- Régression logistique binaire : Dans un ce type, une variable dépendante n'aura que deux types possibles, soit 1 et 0. Par exemple, ces variables peuvent représenter un succès ou un échec, oui ou non , etc

.Régression logistique multinomiale : Dans ce type, la variable dépendante peut avoir 3 types ou plus, Par exemple, ces variables peuvent représenter « Type A » ou « Type B » ou « Type C », dans notre cas c'est le type de la maladie

- Régression logistique ordinale : Dans ce genre, la variable dépendante peut avoir 3 types possibles ou plus, ou les types ayant une signification quantitative. Par exemple, ces variables peuvent représenter « mauvais » ou « bon », « très bon », « excellent » et chaque catégorie peut avoir des scores tels que 0,1,2,3.

III.4.6 Naïve Bayes

L'apprentissage bayésien permet une prédiction basée sur la probabilité. L'algorithme naïf de Bayes est basé sur le théorème de Bayes et suppose une indépendance complète variable. Il s'agit d'un algorithme d'apprentissage supervisé de type classification. C'est Principalement utilisé pour les problèmes de classification de texte.

La formule bayésienne est le résultat des travaux du pasteur Thomas Bayes. Il est basé sur les Probabilités conditionnelle qui peut se traduire par la probabilité qu'un événement se produise sachant qu'un autre événement c'est déjà produit. Cette formule est définie par la relation suivante :

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \qquad \text{(III.3)}$$

Ici (A |B) désigne la probabilité a posteriori de A sachant B,

P(B │ A) désigne la probabilité a posteriori de B sachant A,

P (A) est la probabilité a priori de A ou probabilité marginale de A,

P(B) est la probabilité a priori de B ou probabilité marginale de B

Chapitre IV : Expérimentation & Résultats

IV.1 Introduction

Le botnet Mirai, ce logiciel malveillant capable de transformer des ordinateurs en bots contrôlés à distance est à l'origine d'une cyber-attaque à très grande échelle. Un ou plusieurs botnets Mirai ont été utilisés pour lancer l'une des plus importantes attaques DDoS.

Dans ce chapitre, nous décrivons le principe de notre travail qui consiste à utiliser l'apprentissage numérique pour la détection des botnets. Différents algorithmes d'apprentissage numérique ont été utilisés sur l'ensemble de données Botnet (mirai).La comparaison entre les résultats obtenu nous a permet de sélectionner un meilleur algorithme pour notre modèle. L'ensemble de données a été divisé en deux parties (apprentissage et tests). Avec l'analyse manuelle nous avons utilisé la technique de sélection et d'extraction de caractéristiques pour sélectionner caractéristiques appropriées basées sur la meilleure précision.

IV.2 Démarches

L'idée de base de notre travail est de sélectionner le meilleur algorithme d'apprentissage numérique permettant de mieux détecter les trafics provenant de source risque et atténuer les attaques par déni de service distribué (DDoS) basées sur les botnets dans le réseau IoT.

Nous avons appliqué 4 algorithmes d'apprentissages numériques (Naive Bayes ,SVM, ,KNN, Randomfoerst) sur une base de données constitué des données capturée dans des réseaux d'internet des objets et qui sont libellés par experts comme botnet ou non .

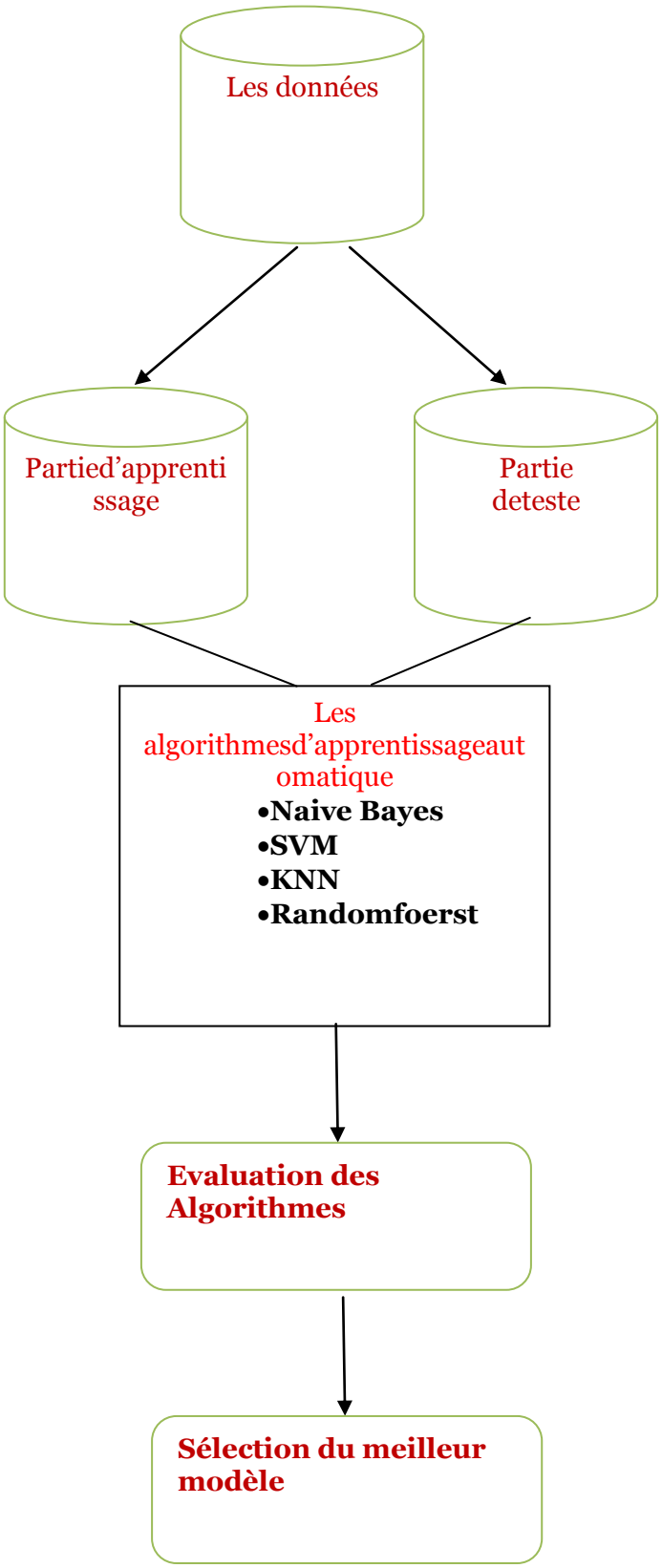
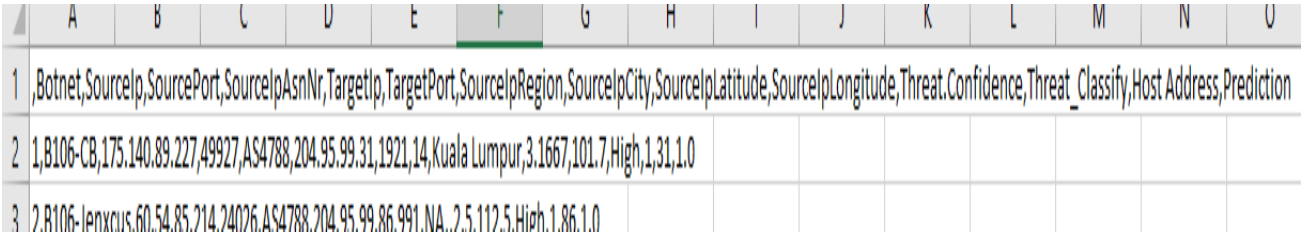


Figure4.1démarches de la séléction

IV.3 Les données utilisées :

Notre base a été récupéré par application de captures, elle regroupe des attaques de type mirai Elle contient en environ de 1750 entrées.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	,Botnet,SourceIp,SourcePort,SourceIpAsnNr,TargetIp,TargetPort,SourceIpRegion,SourceIpCity,SourceIpLatitude,SourceIpLongitude,Threat.Confidence,Threat_Classify,Host Address,Prediction														
2	1,B106-CB,175.140.89.227,49927,AS4788,204.95.99.31,1921,14,Kuala Lumpur,3.1667,101.7,High,1,31,1.0														
3	2,B106-Jenxcus,60.54.85.214,24026,AS4788,204.95.99.86,991,NA,,2.5,112.5,High,1,86,1.0														

Figure 4.2 Extrait du Données

IV.4 Description

1.Serial number

2.Botnet ID

3.Source IpAdresses

4.Source port Adresses

5.Source Ip ASN number

6.Target Ipadress

7.Target port Adress

8.Source IpRegion

9.Source Ip City

10.Source Ip Latitude

11.Source Ip Longitude

12.Threat Confidence

13.Threat_Classify (1 :high , 0 :low)

14.Host

IV.5 Equipements

Le matériel joue un rôle essentiel dans les performances du modèle. Le système que nous avons utilisé tout au long du processus de création du modèle et les tests sont constants. Nous avons utilisé un ordinateur portable Acer sous Windows 10, Système d'exploitation 64 bits avec 8 Go de RAM. Le processeur est core i5 avec une vitesse d'horloge de 1,83 GHz.

IV.5.1 Langage de programmation

IV.5.1.1 Présentation de python

Python est un langage de programmation (au même titre que le C, C++, fortran, java . .)

Développé en 1989. Ses principales caractéristiques sont les suivantes :

- « open-source» : son utilisation est gratuite et les fichiers sources sont disponibles et modifiables
- Simple et trèslisible
- Doté d'une bibliothèque de base très fournie
- Importante quantité de bibliothèques disponibles : pour le calcul scientifique, les statistiques, les bases de données, la visualisation. .
- Grande portabilité : indépendant vis à vis du système d'exploitation (linux, windows, MacOS)
- Orienté objet
- Typage dynamique : le typage (association à une variable de son type et allocation zone mémoire en conséquence) est fait automatiquement lors de l'exécution du programme, ce qui permet une grande flexibilité et rapidité de programmation, mais qui se paye par une surconsommation de mémoire et une perte de performance ;
- Présente un support pour l'intégration d'autres langages.

Il existe deux versions de Python : 2.7 et 3.3. La version 3.3 n'est pas une simple amélioration de la version 2.2. Attention, toutes les librairies Python n'ont pas effectué la migration de 2.7 à 3.3.

IV.5.1.2 Bibliothèques utilisées

Pour traiter l'ensemble de données et mettre en œuvre l'apprentissage automatique, nous avons utilisé de nombreuses bibliothèques python.

- **Sklearn** : la bibliothèque Sklearn est principalement utilisée pour créer la matrice de confusion, pour construire des modèles d'apprentissage automatique, pour diviser un ensemble, pour effectuer le prétraitement des données et pour la procédure d'ingénierie des fonctionnalités.
- **Matplotlib** : la bibliothèque Matplotlib est utilisée pour visualiser les données sous format graphique. Cette bibliothèque prend en charge le graphique à barres, le nuage de points et de nombreux autres graphiques qui aident à comprendre et analyser clairement les résultats obtenus.
- **Pandas :** la bibliothèque Pandas prend en charge l'analyse des données. Nous utilisons la bibliothèque pandas pour importer l'ensemble de données au format de fichier .CSV et pour manipuler les données.
-

IV.6 L'évaluation des algorithmes

L'évaluation des performances du modèle d'apprentissage automatique est effectuée en générant une matrice de confusion pour chaque algorithme de machine Learning automatique à gagner aperçu du type d'erreur commise par l'apprentissage automatique modèle qui nous aide à comprendre les autres métriques telles que précision qui en découlent. Nous avons dérivé l'exactitude, la précision, le rappel et le score F1 de la matrice de confusion pour évaluer les performances du modèle.

True positives (TPs)	False positives (FPs)
False Negatives (FNs)	TrueNegatives (TNs)

Tableau 4.1: matrice de confusion

- **Accuracy :** Accuracy est le rapport entre le nombre de classes correctement prédites et le nombre total de prédictions. Il est présenté en pourcentage. La précision est analysée lorsque les vrais positifs (TP) et les vrais négatifs (TN) sont cruciaux

  $$\text{Accuracy} = ((TPs+TNs)/(TPs+TNs+FPs+FNs)*100)\% \quad (1)$$

- **Précision** : La précision est le rapport entre le nombre de correctement positifs prédits avecle total des positifs prédits. Précisionestanalysé pour minimiser les faux positifs.

  $$\text{Precision} = (TPs/(TPs+FPs)*100)\% \quad (2)$$

- **Recall** : recall(Rappel) est le rapport du nombre de correctement prédit positifs à tous les exemples positifs. Le rappel estanalysé pour minimiser les faux négatifs.

  $$\text{Recall} = (TPs/(TPs+FNs)*100)\% \quad (3)$$

- **F1-Score** : La métrique F-Score combine précision et recall, elle donne une valeur de score unique pour équilibrer à la fois les préoccupations de précision et de rappel. F1-Score est analysé lorsqu'il est faux les positifs et les faux négatifs sont importants.

  $$\text{F1-Score} = ((2*\text{precision}*\text{Recall})/(\text{Precision}+\text{Recall})*100)\% \quad (4)$$

- **Courbe ROC**

  Une courbe ROC (receiver operating characteristic) est un graphique représentant les performances d'un modèle de classification pour tous les seuils

de classification. Cette courbe trace le taux de vrais positifs en fonction du taux de faux positifs :

- Taux de vraispositifs
- Taux de faux positifs

Le taux de vrais positifs (TVP) est l'équivalent du rappel. Il est donc défini comme suit :

$$TVP=VP/(VP+FN)$$

Le taux de faux positifs (TFP) est défini comme suit :

$$TFP=FP/(FP+VN)$$

➢**AUC : aire sous la courbe ROC**

L'AUC représente la probabilité pour qu'un exemple positif aléatoire soit placé à droite d'un exemple négatif aléatoire.

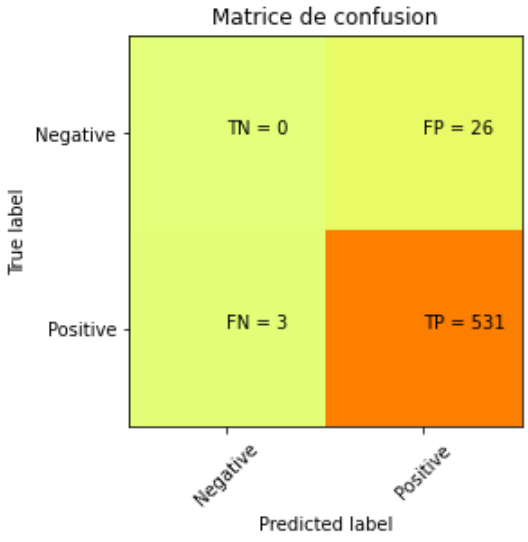
Les valeurs d'AUC sont comprises dans une plage de 0 à 1. Un modèle dont 100 % des prédictions sont erronées a un AUC de 0,0. Si toutes ses prédictions sont correctes, son AUC est de 1,0.

IV.7 Expérimentations et discussions

L'ensemble de données a été divisé en apprentissage et test pour évaluer les performances du modèle. 75% des données ont été utilisées pour l'apprentissage et 25% pour le test.

Nous avons appliqué, les différents algorithmes de machine Learning sur le data set mirai et nous avons obtenus les résultats suivants :

Gaussian Naïve bayes:



precision recall f1-score

0	0.00	0.00	0.00
1	0.95	0.99	0.97
accuracy			0.95

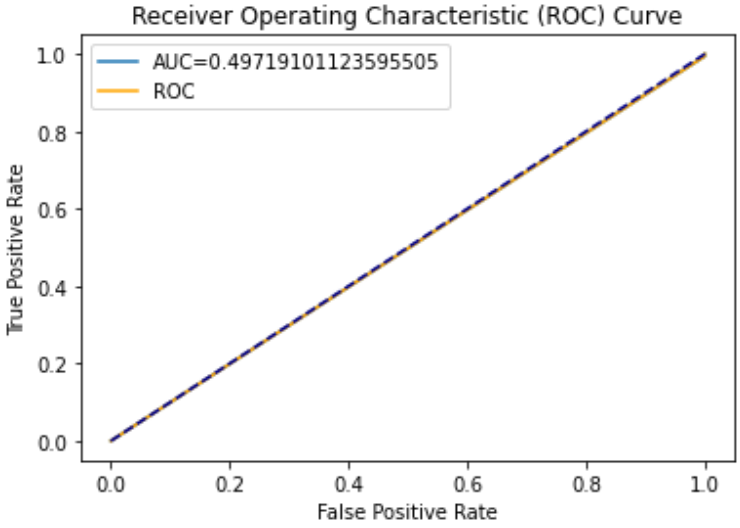
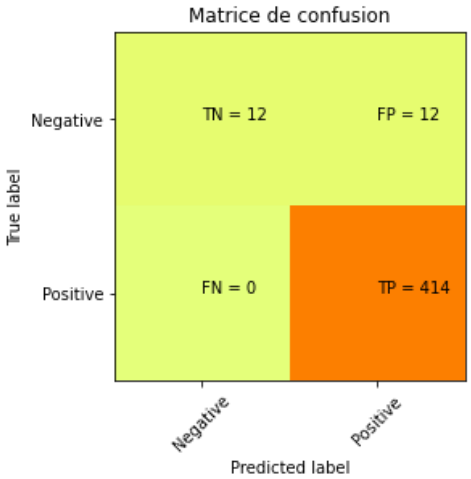


Figure4.3résultats du modèle Naïve bayes

Comme le montre la figure (fig 4.5), pour un jeu de données BoT-IoT réel, avec des bayes Gaussian Naïve algorithme, nous avons observé près de 0.95% d'accuracy mais 0.49% ROC-AUC, et valeur elevée du rappel et du score f1. Cela indique que l'algorithme gaussien n'est pas efficace pour distinguer le botnet et le trafic normal.

1.2Le modèle KNN



	precision	recall	f1-score

0	1.00	0.50	0.67
1	0.97	1.00	0.99
accuracy			0.97

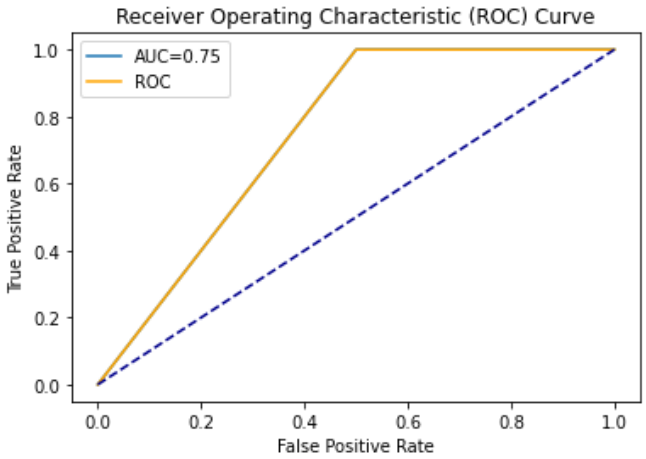
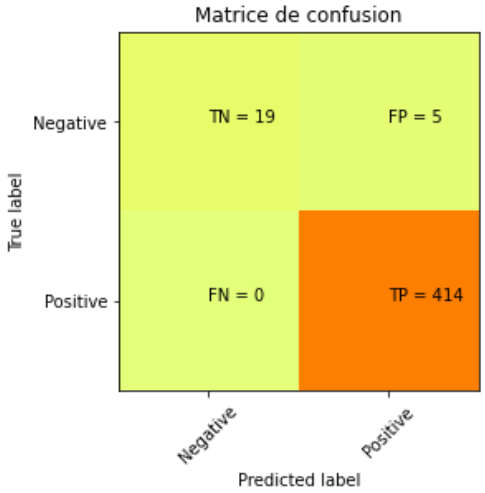


Figure4.4résultats du modèle KNN

L'accuracy et le ROC AUC que nous obtenons sont respectivement de 97,0% et 75%. Pour l'ensemble de données BoT-IoT, le résultat de précision de KNN est bon. Cela indique que le classificateur KNN est un algorithme peu efficace dans système de détection de botnet.

1.3Random_Forest



	precision	recall	f1-score
0	1.00	0.79	0.88
1	0.99	1.00	0.99
accuracy			0.99

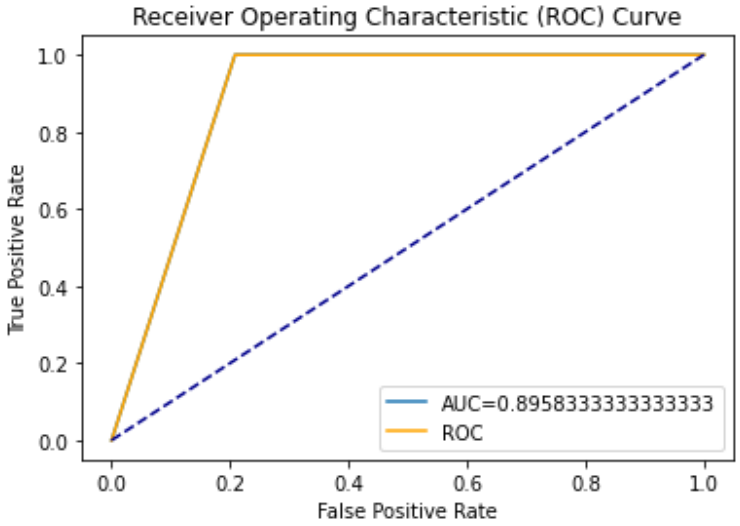
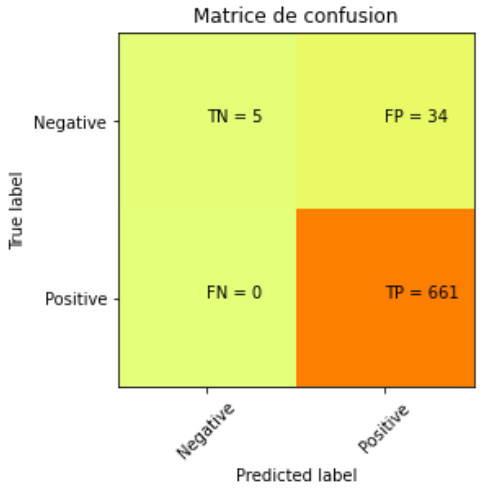


Figure4.5 résultats du modèle Random-Forest

Le modèle Random-Forest a donné un bon résultat d'accuracy de 99% mais un roc-auc de 0.895, les résultats F-mesure ; precision et recall montrent que le système peut détecter les trafic botnet.

1.4Le modèle SVM



	precision	recall	f1-score
0	1.00	0.13	0.23
1	0.95	1.00	0.97
accuracy			0.95

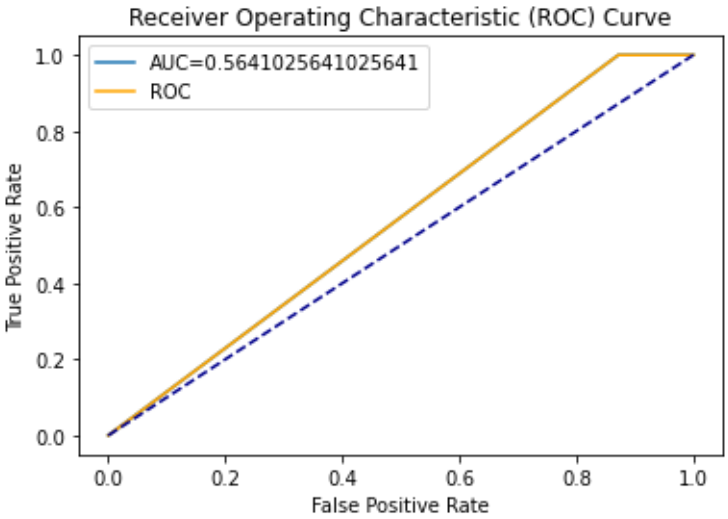


Figure4.6 résultats du modèle SVM

Le modèle présente une accuracy de 95% mais un AUC de 56%. Cela montre que SVM n'est pas assez efficace.

IV.8 Discussion des résultats :

Algorithme	Accuracy	Roc_Auc
Naive Bayes	0.945	0.497
SVM	0.954	0.564
KNN	0.972	0.750
Randomfoerst	0.988	0.895

Tableau4.2 : résultats des expérimentations

Le classificateur Randomforestdonne les meilleurs résultats avec une Accuracy qui égale a 98.8%. KNN a donné comme résultats 95.4%, Naive-Bayes a donné un pourcentage de 94.5 %.

Le random-forest s'avere bon et stable, ce qui nous guide a le considéré comme le meilleur modèle qui peut être implémenté dans un équipement IoT pour atténuer les attaque DDos.

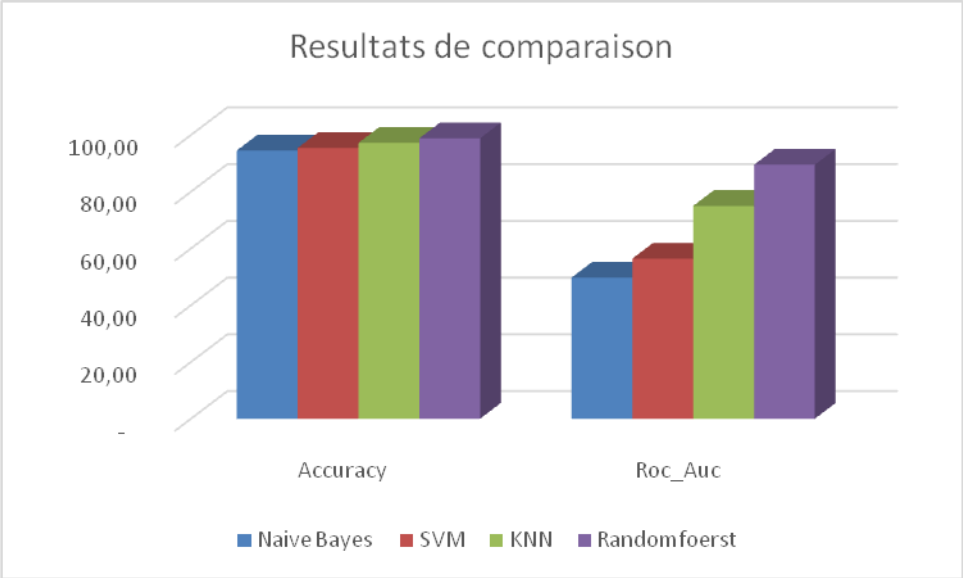


Figure 4.7 Comparaison des performances des algorithmes

IV.9 Conclusion :

Dans ce chapitre, nous avons présenté notre phase d'expérimentations et résultats, nous avons appliqué l'ensemble de 4 algorithmes de machine Learning sur le data-set Mirai, et nous avons calculé les mesures qui nous permettent de choisir le meilleur algorithme, la comparaison a montré que randomforest est le meilleur algorithme qui mieux classifier un trafic s'il est botnet ou non.

Conclusion générale

Conclusion Générale

Dans ce travail, Nous avons abordé la sécurité dans l'Internet des objets, nous avons étudié la possibilité d'appliquer l'intelligence artificielle pour remédier au problème de Botnet ou atténuer les attaques. Nous avons utilisé l'apprentissage automatique.

Les expérimentations réalisées et les résultats obtenus ont montré que l'algorithme randomforest peut fournir une aide proactive au système. A cet égard, il est aisé de l'intégrer sur une passerelle ou un Routeur, et regardez le trafic de flux réseau, et d'autoriser ou bloquer un tel trafic selon le résultat de classification (trafic botnet ou normal).

Nous pouvons affirmer que les approches basées et l'apprentissage automatique peuvent avoir une contribution importante à la détection des structures de botsmalveillants sur un réseau à grande échelle

Références Bibilographies

Références bibliographiques

Références :

[1] (en) Dirk Helbing et EvangelosPournaras (2015) Share/bookmark « Society: Build digital democracy » [archive] [PDF], *Nature*, 2 novembre 2015.
[2] Noto La Diega 2015.

[3] Présentation générale de l'Internet des objets [archive] (rapport ITU-T Y.2060),ITU-T Study Group 20, ITU,juin 2012

[4] « *ITU-T définition de l'Internet des Objets* » [archive du 12 août 2013].

[5] Revenir plus haut en :a et b [PDF] Étude L'Internet des objets [archive], par Pierre-Jean Benghozi et Sylvain Bureau (Pôle de recherche en Économie et Gestion de l'École Polytechnique) et Françoise Massit-Folléa (programme Vox Internet II).

[6] (en) Anonyme. 2008. Internet of Things in 2020. Roadmap for the Future, 1.1 ed.: 27: Infso D.4 Networked Enterprise & RFID; Infso G.2 Micro &Nanosystems in co-operation with the working group RFID of the EPOSS. p. 4

[7] S. Le Pallec, http://2005.jres.org/paper/70.pdf [archive]

[8]Proulx, S. (2005). *Penser les usages des technologies de l'information et de la communication aujourd'hui : enjeux–modèles–tendances [archive]*, ss la dir. de. Lise Vieira et Nathalie Pinède, Enjeux et usages des TIC : aspects sociaux et culturels, Tome, 1, 7-20.
[9] « *The "Only" Coke Machine on the Internet* » [archive]

[10]Ogor, P. (2001). *Une architecture générique pour la supervision sûre à distance de machines de production avec Internet* (Doctoral dissertation, Brest)

[11] Benghozi, P. J., Bureau, S., & Massit-Folea, F. (2008). *des objets. Quels enjeux pour les Européens ? [archive]*
[12] P. Gautier, Objets « connectés », objets « communicants »… ou objets « acteurs », http://www.refondation.org/blog/2385/internet-des-objets-objets-connectes-objets-communicants-ou-objets-acteurs [archive].

[13]Gautier et Gonzalez 2011.
[14] P. Gautier, Internet des objets et perspectives économiques, http://www.i-o-t.org/post/Internet-des-objets-et-perspectives-economiques [archive]

Références bibliographiques

[15]Brousseau, É. (2001). *Régulation de l'Internet* (Vol. 52, No. 7, p. 349-377). Presses de Sciences Po.

[16]Benhamou, Bernard (2009), *L'Internet des objets ; Défis technologiques, économiques et politiques [archive]*, Revue Esprit ; 2009/3 (mars/avril), 270 pages, SBN : 9782909210759; DOI : 10.3917/espri.0903.0137

[17]Benhamou B, Weill M Quelle gouvernance pour l'Internet des Objets ? http://mathieuweill.fr/images/Objets.pdf [archive]

[18] IoTPlanet : Grenoble accueille le 1er forum international dédié aux objets connectés [archive], Channelnews.fr, 2015-10-13

[19]MELITI Nedjema. Architecture Basée Agents pour le diagnostic d‟un système d‟IoT (Internet of Things). Mémoire de Master académique , Architectures Distribuées. Oum-El-Bouaghi : Université Larbi Ben M‟hidi Oum-El-Bouaghi, 2017, 111 p.

[20]Abdmeziem, M.R.; Tandjaoui, D.; Romdhani, I. Architecting the Internet of Things: State of the art. Sens. Clouds 2015, 36, 55–75. [CrossRef]

[21] Al-Fuqaha, A.I.; Guizani, M.; Mohammadi, M.; Aledhari, M.; Ayyash, M. Internet of Things: A survey on enabling technologies, protocols, and applications. IEEE Commun. Surv. Tutor. 2015, 17, 2347–2376. [CrossRef]

[22]Colakovi´c, A.; Hadžiali´c, M. Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues. ˇ Comput.Netw. 2018, 144, 17–39. [CrossRef]

[23]Abdmeziem, M.R.; Tandjaoui, D.; Romdhani, I. Architecting the Internet of Things: State of the art. Sens. Clouds 2015, 36, 55–75. [CrossRef]

[24].Al-Fuqaha, A.I.; Guizani, M.; Mohammadi, M.; Aledhari, M.; Ayyash, M. Internet of Things: A survey on enabling technologies, protocols, and applications. IEEE Commun. Surv. Tutor. 2015, 17, 2347–2376. [CrossRef]

[25]Taivalsaari, A.; Mikkonen, T.A taxonomy of IoT client architectures. IEEE Softw. 2018, 35, 83–88. [CrossRef]

[26]Mattern, F.; Floerkemeier, C.From the internet of computers to the Internet of Things.Lect. Notes Comput. Sci. 2010, 6462, 242–259. [CrossRef]

[27]Abdmeziem, M.R.; Tandjaoui, D.; Romdhani, I. Architecting the Internet of Things: State of the art. Sens. Clouds 2015, 36, 55–75. [CrossRef]

[28] Sharma, C.; Gondhi, N.K. Communication protocol stack for constrained IoT systems. In

Proceedings of the 2018 3rd International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Bhimtal, India, 23–24 February 2018.
[29]Colakovi´c, A.; Hadžiali´c, M. Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues. ˇ Comput.Netw. 2018, 144, 17–39. [CrossRef]
[30] SETHI, Pallavi et SARANGI, Smruti R. Internet of things: architectures, protocols, and applications. Journal of Electrical and Computer Engineering, 2017, vol. 2017.
[31] http://www.futura-sciences.com/magazines/high-tech/infos/dico/d/hightech-rfid-4187
[32 ,33 ,34 ,35 ,36] Internet of Things (IoT) Protocols You Need to Know About" http://www.rs-online.com/designspark/electronics/knowledgeitem/eleven-internet-of-things-iot-protocols-you-need-to-know-about
[37] KHANNA, Abhishek et KAUR, Sanmeet. Evolution of Internet of Things (IoT) and itssignificant impact in the field of Precision Agriculture. Computers andelectronics in agriculture, 2019, vol. 157, p. 218-231.
[38] L'Internet des Objets : 101, Publié par Sameh Ben Fredj http://blog.xebia.fr/2015/12/02/linternet-des-objets-101/
[39] HomeKit. HomeKit - Apple Developer, 2018. URL https://developer.apple.com/ homekit/.
[40] L. Martirano. A smart lighting control to save energy. In Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems, volume 1, pages 132–138, September 2011. doi : 10.1109/IDAACS.2011. 6072726.
[41] Insight Robotics. Forestry-Focused Risk Management, 2018.URL https://www. insightrobotics.com/en/.
[42] Hikob. Systèmes d'acquisition de données stationnaires par Hikob, 2018. URL https://www.hikob.com/instant/
[43] All Traffic Solutions. All Traffic Solutions IoT Solutions for Smart Parking & Transportation Mgmt., 2018.URL http://www.alltrafficsolutions.com/.
[44] D. Gachet, M. de Buenaga, F. Aparicio, and V. Padrón. Integrating Internet of Things and Cloud Computing for Health Services Provisioning : The Virtual Cloud Carer Project. In 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pages 918–921, July 2012. doi : 10.1109/IMIS.2012.25.
[45] Yanzi. Yanzi, 2018.URL https://www.yanzi.se/.[263] Xiaojing Ye and Junwei Huang.A framework for Cloud-based Smart Home. In Proceedings of 2011 International Conference on Computer Science and Network Technology, volume 2, pages 894–897, December 2011. doi :

10.1109/ICCSNT.2011. 6182105.

[46] OnFarm. Onfarm, 2018.URL http://www.onfarm.com/.

[47] AirCasting. AirCasting, 2018.URL http://aircasting.org/.

[48] Alothman, B.Similarity based instance transfer learning for botnet detection.*Int.J. Intell.Comput.Res.(IJICR)* **2018**, *9*, 880–889.[CrossRef]

[49] Morse,A.*Investigation:WannaCry CyberAttack and the NHS*;ReportbytheNationalAuditOffice.Accessed;NationalAuditOffice:London,UK,2018;Volume1.

[50] Amini,P.;Araghizadeh,M.A.;Azmi,R.AsurveyonBotnet:Classification,detectionanddefense. InProceedingsofthe2015InternationalElectronicsSymposium(IES),Surabaya,Indonesia, 29–30September2015;pp.233–238.

[51] Alieyan,K.;ALmomani,A.;Manasrah,A.;Kadhum,M.M.Asurveyofbotnetdetectionbasedon DNS.*NeuralComput.Appl.* **2017**,*28*,1541–1558.[CrossRef]

[52] Li,X.;Wang,J.;Zhang,X.BotnetdetectiontechnologybasedonDNS.*FutureInternet***2017**,*9*,55.[CrossRef]

[53] Singh, M.; Singh, M.; Kaur, S. Issues and challenges in DNS based botnet detection: A survey. *Comput.Secur*.**2019**, *86*, 28–52.[CrossRef]

[54] Gaonkar,S.;Dessai,N.F.;Costa,J.;Borkar,A.;Aswale,S.;Shetgaonkar,P.Asurveyonbotnetdetectiontechniques.InProceedingsof the2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India,24–25February2020;pp.1–6.

[55] Vishwakarma,R.;Jain,A.K.AsurveyofDDoSattackingtechniquesanddefencemechanismsintheIoTnetwork.*Telecommun.Syst.* **2020**,*73*,3–25.[CrossRef]

[56] Hadianto,R.;Purboyo,T.W.Asurveypaperonbotnetattacksanddefensesinsoftwaredefinednetworking.*Int.J.Appl.Eng.Res.*

2018,*13*,483–489.

[57] Prononciation en anglais américain retranscrite selon la norme API

[58] https://en.wikipedia.org/wiki/Denial-of-service_attack

[59]. Alauthman, M. Une approche efficace de la détection de bots en ligne basée sur une technique d'apprentissage par renforcement. doctorat Thèse, Université de Northumbria, Newcastle upon Tyne, Royaume-Uni, 2016

[60]. Limarunothai, R. ; Munlin, MA Tendances et défis des architectures de botnet et des techniques de détection. J. Inf. Sci. Technol. **2015** ,
5, 51-57.

[61]. Ghafir, I. ; Svoboda, J.; Prenosil, V. Une enquête sur la détection du trafic de commande et de contrôle des botnets. Int. J. Adv. Calcul. Réseau Sécurisé.
2015 , 5, 7580. Ateliers, Victoria, C.-B., Canada, 13-16 mai 2014 ; p. 7-12.

[62]. Miller, S.; Busby-Earle, C. Le rôle de l'apprentissage automatique dans la détection des botnets. Dans Actes de la 11e Internationale 2016
Conférence sur la technologie Internet et les transactions sécurisées (ICITST), Barcelone, Espagne, 5-7 décembre 2016 ; p. 359-364.

[63]. Vania, J. ; Meniya, A.; Jethva, H. Une revue sur les botnets et la technique de

[64]. Asha, S.; Harsha, T.; Soniya, B. Analyse sur les techniques de détection de botnet. Dans Actes de la Conférence internationale 2016 sur
Les progrès de la recherche dans les systèmes de navigation intégrés (RAINS), Karnataka, Inde, 6-7 mai 2016 ; p. 1–4.

[65] « DDoS ATTACK HANDBOOK Service Providers » , (2018), Allot Communications, Ltd.

[66] B. Krebs. (2016) KrebsOnSecurity Hit With Record DDoS.[Online]. Available: https://krebsonsecurity.com/2016/09/krebsonsecurityhit-with-record-ddos (consulter le 20 juin 2021)

[67]« Botnets Architectures, Countermeasures, and Challenges. byAnagnostopoulos, MariosKambourakis, GeorgiosMeng, Weizhi Zhou, Peng (z-lib.org).pdf ».

[68]. Alothman, B. Similarity based instance transfer learning for botnet detection. Int. J. Intell. Comput. Res. (IJICR) 2018, 9, 880–889. [CrossRef]

[69]. Silva, S.S.; Silva, R.M.; Pinto, R.C.; Salles, R.M. Botnets: A survey. Comput.Netw. 2013, 57, 378–403. [CrossRef]

[70]. Ghafir, I.; Svoboda, J.; Prenosil, V. A survey on botnet command and control traffic detection. Int. J. Adv. Comput. Netw.Secur. 2015, 5, 7580

[71]. Miller, S.; Busby-Earle, C. The role of machine learning in botnet detection. In Proceedings of the 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST), Barcelona, Spain, 5–7 December 2016; pp. 359–364.

[72]. Limarunothai, R.; Munlin, M.A. Trends and challenges of botnet architectures and detection techniques. J. Inf. Sci. Technol. 2015, 5, 51–57.

[73]Silva, SS; Silva, RM; Pinto, RC; Salles, RM Botnets : Une enquête. *Calcul. Réseau* **2013** , *57* , 378-403. [CrossRef]

[74] Ghafir,I.;Svoboda,J.;Prenosil,V.Asurveyonbotnetcommandandcontroltrafficdetection.*Int.J.Adv.Comput.Netw.Secur.* **2015**,*5*,7580.

[75] Vania,J.;Meniya,A.;Jethva,H.Areviewonbotnetanddetectiontechnique.*Int.J.Comput.TrendsTechnol.***2013**,*4*,23–29.

[76] Asha,S.;Harsha,T.;Soniya,B.Analysisonbotnetdetectiontechniques.InProceedingsofthe2016InternationalConferenceonResearchAdvancesinIntegratedNavigationSystems(RAINS),Karnataka,India,6–7May2016;pp.1–4.

[77] «Botnet Business ». https://securelist.fr/botnet-business/59191/ (consuter le 2 avril 2021)

[79] « apprentissage automatique » [archive], Le Grand Dictionnaire terminologique, Office québécois de la langue française (consulté le28 janvier 2020).

[80]↑ Revenir plus haut en :a et b Commission d'enrichissement de la langue française, « Vocabulaire de l'intelligence artificielle (liste de termes, expressions et définitions adoptés) », Journal officiel de la République française no 0285 du 9 décembre 2018 [lire en ligne [archive]] [PDF].

[81] A. Géron, « Machine Learning avec Scikit-Learn », Dunod : Paris, 2017.

[82] M. Taffar, « Initiation à l'apprentissage automatique », Cours Master, Ment Informatique Faculté des Sciences Exactes et de l'Informatique.

[83] Y. Benzaki, « Machine Learning made easy », https://mrmint.fr/, 2019.

[84]A. Géron, « *Machine Learning avec Scikit-Learn* », Dunod : Paris,2017.

[85]Y. Benzaki, « *Machine Learning made easy* », https://mrmint.fr/,2019.
[86]J. B. Metomo, « *Machine Learning : Introduction à l'apprentissage automatique* »

https://www.supinfo.com/articles/single/6041-machine-learning-introduction-

apprentissage- automatique, 10 octobre2017.

[87] B. E. Boser, I. M. Guyon, et V. N. Vapnik, "A training algorithm for optimal margin classifiers," Dans Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, pp. 144-152.
[88]Sanskruti ,P.(May. 2018).ApplyingSupervised Machine Learning Algorithms for Analytics of SensorData.IOSR Journal of Engineering (IOSRJEN) .Vol. 08, ||V (IV) || PP 16-22
[90] Arbre de décision https://www.lovelyanalytics.com/2016/08/16/decision-tree-comment-ca-marche/ ,2016
[91] Moutarde, F. (2008). Brève introduction aux arbres de décision. Paris, Centre de Robotique (CA OR), École des Mines de Paris.
[92] L. Breiman, "Randomforests," Machine learning, vol. 45, no. 1, pp. 5–32, 2001]
[93]Lydia DEKKICHE, ' Classification des arythmies ECG avec des méthodes de Machine Learning et de Deep Learning' ,mémoire de fin d'etude pour obtention du diplôme master en informatique , université de mouloud mammeri de tiziouzou ,2020
[94] *KavyaS ,Sarath T S , Siddharth M , Abhitha S and AkhilKuriakose « 'A STUDY OF DDOS ATTACK BY BOTNET INFECTED IOT DEVICES USING MULTIPLE MACHINE LEARNING CLASSIFIERS' », Cochin University of Science and Technology, November 2019.*
[95] *David G.Kleinbaum, MitchelKlein.« Introduction to LogisticRegression », 2010 :*