République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Chadli Bendjedid
Faculté des Sciences et de la Technologie
Département d'Informatique

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جامعة الشاذلي بن جديد
كلية العلوم و التكنولوجيا
قسم الاعلام الالي

جامعة الشاذلي بن جديد
UNIVERSITE CHADLI BENDJEDID

MEMOIRE

Présenté par

ABES Nouh

Pour l'obtention de diplôme de

MASTER

Filière : Informatique

Spécialité : Systèmes Informatiques Intelligents

Thème

Un système de détection des attaques DDOS basé sur l'apprentissage automatiques

Soutenu le : 30/06/2022

Devant le Jury composé de :

Qualité	Nom et Prénom	Grade	Université
Président	**Mme. Melle Ziani A**	**MCB**	**Chadli Bendjedid El-Tarf**
Encadreur	**Mr. BETOUIL A.A**	**MCB**	**Chadli Bendjedid El-Tarf**
Examinateur	**Mme. Ahmed Malik N**	**MCB**	**Chadli Bendjedid El-Tarf**

Année Universitaire : 2021/2022

Remerciements

Tout d'abord, je remercie Dieu le Tout-Puissant de m'avoir donné la force morale et physique et de m'avoir permis d'accomplir ce travail. Je tiens à remercier mon encadrant Dr. BETOUIL Ali Abdelatif, pour m'avoir dirigé dans ce travail. Je le remercie pour sa disponibilité, son suivi, ses précieux conseils et son aide.

Nous adressons nos sincères remerciements à tous les professeurs, conférenciers, et toutes les personnes qui ont orienté leurs réflexion, écrits, conseils et critiques et ont accepté de nous Rena rencontrer et Co-dansera de répondre aux questions.

Merci à mon père, à ma mère, à ma sœur, à mon frère, pour leur compréhension, leur patience et leur soutien. Alors, à toute ma famille. Ils ont toujours été la source de mon succès, pour cela je ne saurai jamais comment les remercier. Je prie Dieu de me les garder le plus longtemps possible.

Dédicace

Je dédie cet acte humble à tous ceux qui me connaissent, en particulier à mon père et ma mère pour leur sacrifice et leur soutien. Un mon frère Ayman. A mes amis Hacen et Ayman pour leur soutien. À tous mes amis et collègues. Sans oublier tous les professeurs qui m'ont fait passer du primaire, du collège et du lycée au supérieur.

Résumé

Le développement technologique a rendu Internet vulnérable à de nombreuses menaces de sécurité. Cela a conduit à une mauvaise sécurité du système Internet, d'autant plus qu'il existe des menaces d'attaques DDoS.

L'attaque DDoS est l'une des techniques de piratage les plus puissantes. En tant qu'arme principale des pirates, ils utilisent principalement ces types d'attaques pour consommer les ressources du réseau.

Les travaux menés dans le cadre de ce mémoire visent à répondre aux problématiques d'attaques DDoS sur Internet et visent à élaborer une solution de détection intelligente des attaques DDoS via différents modèles de machine learning. Ensuite, nous avons pris l'initiative d'introduire différents concepts de sécurité réseau et nous nous sommes concentrés sur les problèmes et les solutions de ces attaques.

Dans ce contexte, nous utilisons le dernier ensemble de données DDoS pour construire un système de détection d'attaques DDoS, basé sur l'apprentissage automatique où nous avons utilisé plusieurs algorithmes tels que l'arbre de décision, la machine à vecteurs de support (SVM), etc.

Mots-clés : Apprentissage profond, DDOS, SVM, IDS. Apprentissage automatique.

ملخص

جعل التطور التكنولوجي الإنترنت عرضة للعديد من التهديدات الأمنية. أدى إلى ضعف أمان نظام الإنترنت ، خاصة وأن هناك تهديدات بهجمات DDoS.

يعد هجوم DDoS أحد أقوى تقنيات القرصنة. كسلاح أساسي للهاكر ، يستخدمون بشكل أساسي هذه الأنواع من الهجمات لاستهلاك موارد الشبكة.

يهدف هذا العمل المنفذ في إطار هذه الرسالة إلى الاستجابة لمشاكل هجمات DDoS على الإنترنت ويهدف إلى إعداد حل للكشف الذكي عن هجمات DDoS عبر نماذج التعلم الآلي المختلفة. ثم أخذنا زمام المبادرة لتقديم مفاهيم مختلفة لأمن الشبكات وركزنا على مشاكل هذه الهجمات وحلولها.

في هذا السياق، نستخدم أحدث مجموعة بيانات DDoS لبناء نظام اكتشاف هجمات DDoS ، استنادًا إلى التعلم الآلي حيث استخدمنا العديد من الخوارزميات مثل شجرة القرار ، وآلة ناقلات الدعم (SVM) ، وما إلى ذلك.

الكلمات الرئيسية: التعلم العميق, DDOS, SVM, IDS, تعلم الآلة .

Abstract

This technological development has made the Internet vulnerable to many security threats. This has led to poor internet system security, especially since there are threats of DDoS attacks.

DDoS attack is one of the most powerful hacking techniques. As the primary weapon of hackers, they mainly use these types of attacks to consume network resources.

The work carried out within the framework of this letter aims to respond to the problems of DDoS attacks on the Internet and aim to develop a solution for the intelligent detection of DDoS attacks via different machine learning models. Then we took the initiative to introduce different network security concepts and focused on the problems and solutions of these attacks.

In this context, we use the latest DDoS dataset to build a DDoS attack detection system, based on machine learning where we used several algorithms such as decision tree, support vector machine (SVM), etc.

Keywords: Deep learning, DDOS, SVM, IDS. Machine Learning

Table des matières

Introduction Générale 16

Chapitre 1 : Sécurité informatique

1 Introduction 18

2 Informatiques attaques 18

2.1 Attaque de l'homme du milieu..... 18

2.2 Attaques par mot de passe 19

2.2.1 Attaques par force brute 19

2.2.2 Attaques par dictionnaire 19

2.3 Attaques SQL 19

2.4 Logiciels malveillants 20

2.5 Attaques de phishing 21

3 Les attaques DDOS 22

3.1 Introduction 22

3.2 C'est quoi une attaque DDoS? 22

3.3 Les types des attaques DDOS 23

3.3.1 Attaques basées sur le volume..... 24

- L'attaque Amplification DNS 24
- Inondation UDP 25

3.3.2 Attaques de protocole 26

- L'attaque SYN flood 26
- L'attaque Ping of death 27

3.3.3 Attaques d'application 28

- L'attaque Back 28
- L'attaque Slow loris 29

4 Quelques vecteurs d'attaque 29

4.1 Les botnets 29

4.2 Les attaques basées sur la réflexion 29

4.3 Les attaques basées sur l'amplification 30

4.4 Les attaques ciblant des applications 30

5 sécurités informatiques 31

6 Techniques de sécurité 31

7 Conclusion 32

Chapitre 2 : SYSTÈMES DE DÉTECTION D'INTRUSION

1. Introduction 33

2 Définition..... 33

3 Modèle général..... 33

4 types d'ID..... 34

4.1 IDS basé sur l'hôte (HIDS) 34

4.2 IDS basé sur le réseau (NIDS) 35

4.3 IDS hybrids..... 36

5 techniques IDS..... 37

6 Conclusions..... 37

Chapitre 3: Machine Learning

1. Introduction..... 38

2 Machine Learning..... 38

2.1 Définition..... 38

2.2 Types Machine Learning (d'apprentissage automatique) 39

2.2.1 Apprentissage supervisé..... 39

2.2.2 Apprentissage non supervisé..... 41

2.2.3 Apprentissage semi-supervisé..... 41

2.2.4 Apprentissage par renforcement..... 42

3 Algorithmes de Classification..... 43

3.1 Naïve bayes..... 43

3.2 L'algorithme K Nearst Neighbors (K-NN) 43

3.3 Support vecteur Machine(SVM) 44

3.3.1 Définition 44

3.3.2 Comment fonctionnent les SVM ? 45

3.3.3 Une SVM peut-elle permettre de résoudre une régression ? 45

3.3.4 Comment fonctionnent les SVM ? 45

3.4 Random Forest Classifier 45

4 Deep Learning (Apprentissage en profondeur) 46

4.1 Définition..... 46

4.2 Fonctionnement du deep Learning..... 46

4.3 Applications du deep Learning..... 47

5 Métriques utilisés 47

5.1 Confusion matrix 47

5.2 Accuracy 48

5.3 Classification report 48

6 Conclusion 48

Chapitre 4 : Réalisation

1 Introduction..... 49

2 Conception de notre solution 49

3 Choix du data set..... 50

4 Implémentions 51

5 Les outils de développement 52

5.1. Python 52

5.2. Sklearn 53

5.3. Google Colaboratory 53

6. Conclusions..... 54

Chapitre 5 : Évaluation et Discussion

1 Introduction..... 55

2 Résultat des Algorithmes de machine Learning..... 55

2.1 Application de l'algorithme «le classificateur de forêt aléatoire» …… 55

2.2 Application de l'algorithme «Les machines à vecteurs de support» …… 56

3 Comparaison entre les algorithmes …….. 56

4 Comparaison entre nos résultats et les résultats d'un autre article……...… 57

5 Conclusion……. 57

Conclusion générale ……..…….. 58

Liste des figures

Chapitre 1 : Sécurité informatique

Figure 1.1 an example of MITM attack..... 19

Figure 1. 2 Cycle de vie des attaques de phishing basées sur des e-mails de phishing et des sites Web de phishing 22

Figure 1.3 – Exemple d'attaque **DDoS**..... 23

Figure.1 4 – Le modèle OSI 23

Figure 1.5 – Exemple d'amplification 24

Figure.1 6 – Exemple d'attaques udp flood 25

Figure 1.7 – L'architecture du mécanisme d'attaque P2P 25

Figure 1.8 – Exemple d'attaque de protocole..... 26

Figure 1.9 – Les détails du mécanisme d'attaque SYN Flood 27

Figure 1.10 – L'architecture d'attaque Ping of death 27

Figure 1.11 – Exemple d'attaque de la couche d'application 28

Figure 1.12 – l'architecture d'attaque retour 28

Figure1.13 – Les détails du mécanisme d'attaque de Slow loris 29

Chapitre 2 : SYSTÈMES DE DÉTECTION D'INTRUSION

Figure 2. 1 Generic intrusion detection model proposed by the IDWG 34

Figure 2. 2 IDS basé sur l'hôte. 35

Figure 2. 3 IDS basé sur le réseau. 36

Figure 2. 4 IDS hybrids. 36

Chapitre 3: Machine Learning

Figure 3.1 – Un jeu d'entraînement étiqueté pour un apprentissage supervisé. 39

Figure 3.2 – La Classification et la Régression 40

Figure 3.3 – A Gauche : Problème linéairement séparable (Frontière linéaire). A Droite : Problème non linéairement séparable 41

Figure 3.4 Présentation de l'apprentissage non supervisé. Les échantillons d'entrée sont regroupés en grappes en fonction des modèles sous-jacents. …….. 41

Figure 3.5 Aperçu de l'apprentissage semi-supervisé. Les clusters formés par une grande quantité de données non étiquetées sont utilisés pour classer une quantité limitée de données étiquetées…….. ……. .42

Figure 3.6 Vue d'ensemble de l'apprentissage par renforcement. Un agent observe l'état de l'environnement et effectue des actions pour maximiser une récompense globale …….. 42

Figure 3.7 – Random Forest Classifier …… 46

Chapitre 4 : Réalisation

Figure 4.1 – Les étapes de machine Learning …… 49

Figure 4.2 – Le nombres d'attributs …….. 51

Figure 4.3 - Divisez l'ensemble de données en ensemble de formation et de test …… 52

Figure 4. 4 – importation des algorithmes…… 52

Figure 4.5 - Déclaration des algorithmes…… 52

Liste des tableaux

Chapitre 4 : Réalisation

Tableau 4.1 - la description de chaque attribut. 50

Chapitre 5 : Évaluation et Discussion

Tableaux 5.1 – L'algorithme « le classificateur de forêt aléatoire » 55

Tableaux 5.2 – L'algorithme « Les machines à vecteurs de support ». 56

Tableaux 5.3 – Comparaison entre les algorithmes. 56

Tableaux 5.4 – Comparaison entre nos résultats et les résultats d'un autre article. 57

Liste des acronymes

DDoS : Déni de service distribué.

DNS : nom de domaine du service.

UDP : protocole de datagramme utilisateur.

IDS : SYSTÈMES DE DÉTECTION D'INTRUSION.

Introduction Générale

Internet devient de plus en plus populaire aujourd'hui car il offre à un grand nombre d'utilisateurs des services performants et des supports de stockage de données évolutifs.

Malheureusement, l'un des problèmes des écuritécritiques qui menacent l internet est les attaques DDoS qui sont devenues l'une des menaces puissantes , il peut utiliser des centaines de milliers d'hôtes compromis situés dans déférents réseaux dans le monde pour attaquer la cible dans internet en utilisant l'usurpation d'adresse IP, le nombre d'attaques DDoS signalés a considérablement augmenté. La plus grande attaque signalée par un fournisseur de services en 2017 était de 600 Gbit/s. ce qui rend la détection de la source de l'attaque complexe et décille. Il refuse également aux utilisateurs légitimes l'accès aux services fourni par le cloud ou arrête une organisation pour une longue période de temps. En 2016, pendant deux heures une attaque DDoS a été menée sur Amazon, Twitter et Spotify, ce qui a entraîné d'énormes pertes financières a cause de l'interruption de service.

Par conséquent, les mécanismes de défense traditionnel sont devenue inefficace contre ce genre d'attaque c'est pour ça la plupart des études ont recours à des solutions intelligents comme data mining, deep Learning et machine Learning etc., dans ce contexte on va présenté une solution intelligente de détection des attaques DDoS via les differents modèles du machine Learning, car c'est l'un des piliers les plus courants de l'IA, Elle est capable d'analyser des données complexes non structurées et fournit une solution aux tâches avec des règles dicible à définir.

Afin d'obtenir des bons résultats nous suivons les étapes de développement d'un projet machine Learning.

Ce mémoire compose de cinq chapitres organisés de la manière suivante :

Chapitre 1 : Sécurité informatique

Présente la sécurité informatique, ses différentes techniques. Avant cela, le chapitre présente quelques attaques informatiques, puis il décrit l'attaque DDOS, son historique, et son architecture

Chapitre 2 : SYSTÈMES DE DÉTECTION D'INTRUSION

Ce chapitre est consacré à la présentation des systèmes de détection d'intrusion par la description de sa définition, son modèle général, et ses différents types. Enfin, nous présentons les différentes techniques utilisées pour détecter les attaques DDOS.

Chapitre 3 : Machine Learning

Ce troisième chapitre présente le concept de machine Learning, en insistant sur sa définition, ses différents types. Ensuite, le chapitre présente le nouvel axe de recherche « deep Learning » qui a été développé pour relever les défis du Big Data. Enfin, nous présentons les différents types d'algorithmes de classification.

Chapitre 4 : Réalisation

Dans ce chapitre, nous donnons un aperçu de notre solution et présentons notre jeu de données sélectionné, et pour bien comprendre notre travail, nous avons donné les étapes que nous avons suivies dans notre projet et les outils utilisés.

Chapitre 5 : Évaluation et Discussion :

Dans ce chapitre, nous présentons un résumé de notre travail ainsi que les résultats que nous obtenons.

Chapitre 1 : SÉCURITÉ INFORMATIQUE

1 Introduction

Le cyber menace mondiale se développe à un rythme rapide, avec le nombre croissant de violations de données qui se produisent chaque année. Ainsi, l'importance de la sécurité informatique a augmenté. La sécurité informatique exige la confidentialité, l'authentification, l'intégrité et la non-répudiation pour empêcher les pirates d'accéder illégalement aux données.

La sécurité informatique est un ensemble de ressources matérielles et logicielles mises en œuvre pour diminuer la vulnérabilité d'un système face aux menaces telles que les attaques DDoS qui est l'un des risques les plus importants en matière de cyber sécurité aujourd'hui.

Dans ce chapitre, nous allons présenter les différentes attaques informatiques. Ensuite, nous expliquerons les attaques DDOS, son histoire et son architecture. La sécurité informatique est expliquée ci-dessous, avant de présenter ses différentes techniques et de conclure le chapitre.

2 attaques informatiques

2.1 Attaque de l'homme du milieu

Une attaque Man in The Middle est un cyber attaque courante où les pirates relaient, écoutent et manipulent les messages [1] [2] entre deux parties qui pensent que la communication se fait directement entre elles [3]. L'attaque MITM est un tiers malveillant [4].

Les attaquants du MITM ciblent principalement les serveurs [1]. La figure 1.1 montre un exemple d'attaque MITM.

Il existe de nombreux types d'attaques Man in The Middle telles que l'empoisonnement du cache ARP, l'usurpation DNS, le détournement de session [5] et le détournement SSL[6].

Pour détecter les attaques MITM, il existe quelques méthodes : utiliser un logiciel de détection ARP et utiliser des entrées ARP statiques [7].

Steinmetz et al. [8] Ont proposé quatre schémas de détection : détection de la commutation entre les secteurs, détection des altérations de la force du signal reçu, détection de l'intervalle de la balise et détection à l'aide des compteurs de balises. Ces schémas peuvent être utilisés pour la détection d'attaques Man in The Middle.

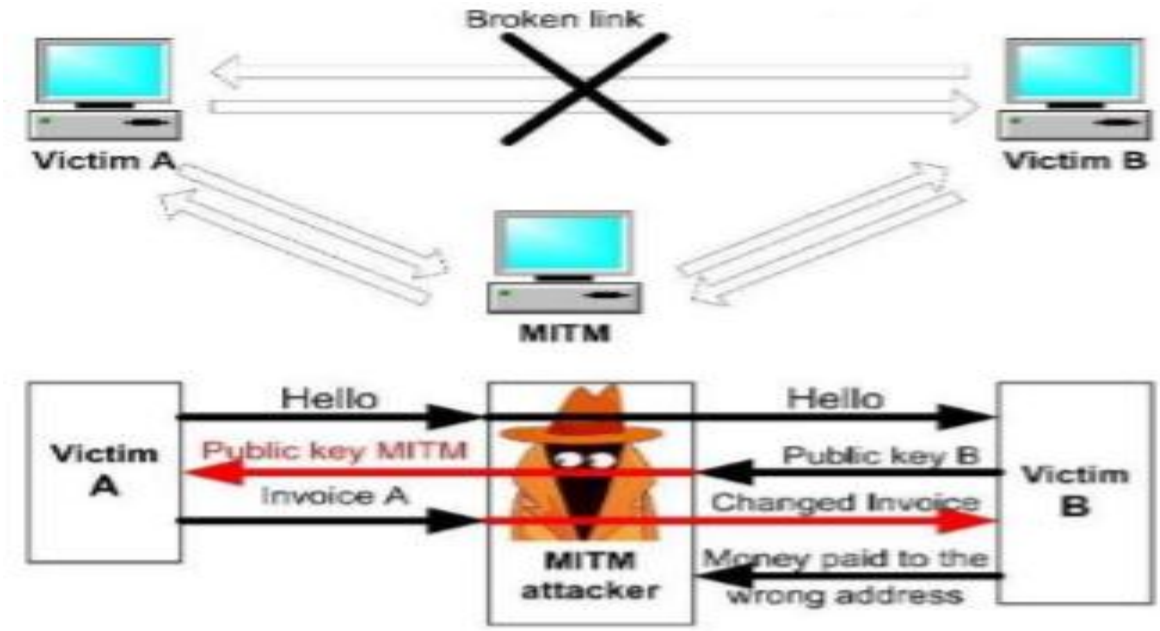


Figure .1.1 Un exemple d'attaque MITM [6].

2.2 Attaques par mot de passe

L'authentification des utilisateurs avec des mots de passe est une méthode [9] bien connue pour accéder à leurs informations sur les sites Web [10].

De nombreuses personnes utilisent un dictionnaire de mots pour créer des mots de passe qui seront stockés dans des bases de données sous forme de valeurs de hachage au lieu de texte brut [10]. Certains utilisateurs choisissent un mot de passe court pour s'en souvenir facilement [11].

En conséquence, les pirates peuvent deviner ces mots de passe en utilisant la force brute, la table arc-en-ciel et les attaques par dictionnaire [12].

2.2.1 Attaques par force brute

Il s'agit d'une attaque où les pirates utilisent toutes les combinaisons possibles de chiffres alpha et d'images graphiques jusqu'à ce que l'on fonctionne [11][12].

2.2.2 Attaques par dictionnaire

Dictionnaire Attac se lance en utilisant une liste possible de mots du dictionnaire. Il donne de bons résultats avec les mots de passe hachés [10].

2.3 Attaques SQL

L'attaque par injection SQL (SQLIA) est l'une des attaques de base de données les plus courantes [13], elle donne un accès illimité aux attaquants pour extraire les données sensibles stockées dans la base de données [14].

Les attaquants utilisent des commandes SQL simples pour restructurer le code SQL et exécuter du code vulnérable dans l'application Web [15].

De nombreuses technologies sont vulnérables aux attaques SQL telles que PHP, JSP, ASP, ASP.net. De plus, les formulaires Web sont également vulnérables à ces types d'attaques telles que les formulaires de connexion, les commentaires, le support client, les paniers d'achat [16].

2.4 Logiciels malveillants

Un logiciel malveillant est un type de logiciel délibérément conçu pour être malveillant et installé à l'insu de l'utilisateur [17]. Il sert à faire des dégâts ou à s'infiltrer dans les systèmes [18].

Il existe de nombreuses caractéristiques des logiciels malveillants, telles que la capacité de réplication, de propagation, d'auto-exécution et de corruption des systèmes informatiques [19].

Il existe deux catégories de techniques de détection : la détection basée sur les anomalies et la détection basée sur les signatures [20].

Les logiciels malveillants ont de nombreux types selon les logiciels malveillants basés sur le réseau (par exemple, les logiciels espions, les logiciels publicitaires, les cookies, les portes dérobées, les chevaux de Troie, les renifleurs, les spam et les bonnet.) et les logiciels malveillants ordinaires (par exemple, les virus, les vers et les bombes logiques.)[17].

Un logiciel espion est un logiciel malveillant qui s'installe dans un ordinateur à l'insu de ses propriétaires pour collecter leurs informations [19].

Adar est un raccourci vers un logiciel financé par la publicité, il diffuse des publicités pour générer un profit financier pour son auteur [19].

Les cookies sont des informations stockées par un navigateur Web sur la machine d'un utilisateur, mais les logiciels espions peuvent les utiliser [19].

Les portes dérobées, ou trappes, sont un code malveillant d'une application ou d'un système, elles sont écrites pour accorder l'accès au système aux programmeurs sans l'exigence d'aucune méthode d'authentification [19] [20].

Les chevaux de Troie sont des logiciels malveillants intégrés dans une application qui semblent remplir certaines fonctions utiles, mais qui effectuent désormais des actions non autorisées [19].

Les renifleurs sont des programmes conçus pour intercepter et enregistrer le trafic sur un réseau [19].

Le spam ou courrier indésirable est un ensemble de logiciels qui envoient en diffusion des messages identiques à de nombreuses cibles [19].

Un virus est un code qui peut se répliquer dans une autre application tout au long de son infection [19].

Un ver est un code logiciel capable de s'auto-répliquer sur une machine cible indépendamment de tout programme. Il n'est pas nécessaire qu'un programme hôte commence le cycle de vie d'un ver. Le ver augmente le taux de propagation du système en créant de nouvelles copies de lui-même [19].

Bomba Logico est un programme qui reste silencieux jusqu'à son activation en fonction de la date et de l'heure [19].

2.5 Attaques de phishing

L'hameçonnage est une attaque d'ingénierie sociale qui vise à exploiter les vulnérabilités des opérations système causées par les utilisateurs du système [21]. Les attaques de phishing commencent par envoyer à l'escroc par phishing un e-mail qui semble provenir d'une véritable organisation, contenant des liens sur lesquels cliquer pourrait soit diriger la victime vers de fausses pages Web où l'utilisateur est invité à fournir ses informations d'identification, soit installer des logiciels espions sur l'appareil. . Le motif de l'attaquant derrière ces escroqueries peut être le vol d'identité, le gain financier ou la notoriété [22].

La figure 1. 3 montres le cycle du phishing.

Plusieurs approches heuristiques telles que les améliorations technologiques, l'ingénierie des processus et la formation des utilisateurs sont suivies pour résoudre le problème du phishing [23].

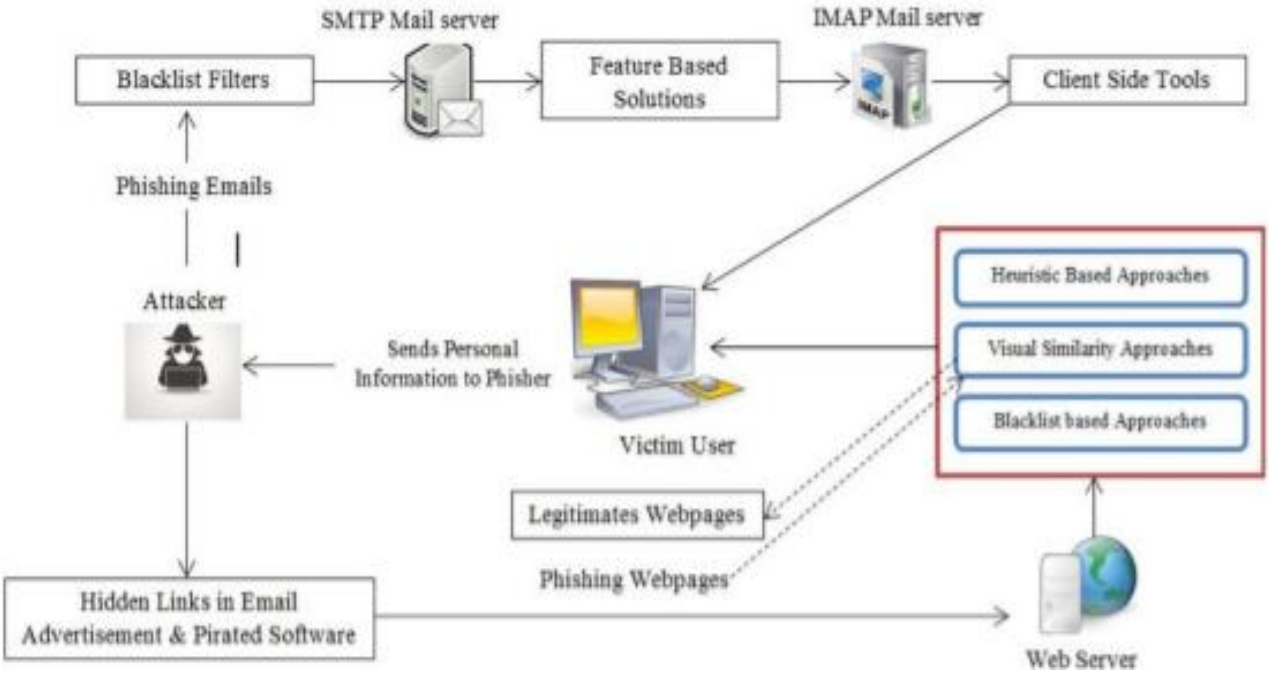


Figure 1. 2 Cycle de vie des attaques de phishing basées sur des e-mails de phishing et des sites Web de phishing [23].

3 Les attaques DDOS

3.1 Introduction

L'attaque Dos est une attaque qui vise à rendre une application informatique incapable de répondre aux requêtes de ses utilisateurs, les auteurs du disent que le déni de service distribué (DDoS) est une attaque Dos émise depuis plusieurs origines distinctes. Ce type d'attaque est extrêmement complexe à bloquer. De plus, les méthodes traditionnelles de la détection et l'atténuation sont tout simplement insu sante par rapport aux déférents méthodes d'attaques DDoS désormais utilisées par les pirates.

3.2 Définition

L'attaque par déni de service distribué, ou DDoS (en anglais Distributed Denial of Service), vise à perturber ou paralyser totalement le fonctionnement d'un serveur informatique en le bombardant à outrance de requêtes erronées.

L'objectif peut être d'éjecter les services en ligne ou le réseau de l'entreprise en saturant l'unedesressourcesdusystème:bandepassante,espacedestockage,puissancedetraitement de la base de données, calculs du processeur, RAM, l'ouverture d'un grand nombre de nouvelles sessions TCP dans un intervalle du temps très court, ou encore d'un nombre trop important de traitements concurrents exécutés par une base de données. Le DoS distribué appelé DDoS est la méthode

[20]d'attaque DoS la plus populaire car il est capable de causer des émets plus graves facilement et rapidement. Voici la figure 1.3 qui montre un exemple simple d'une attaque DDoS.

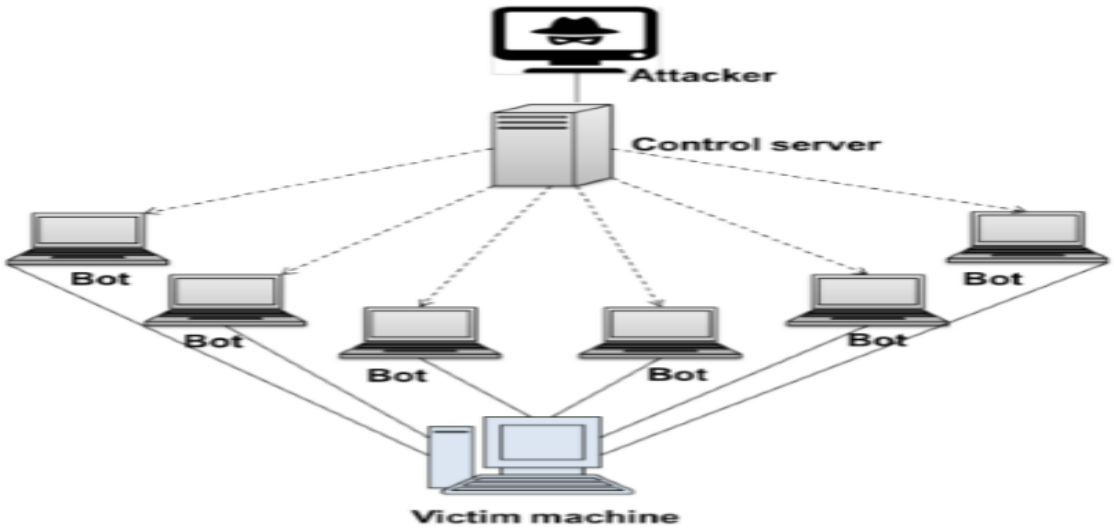
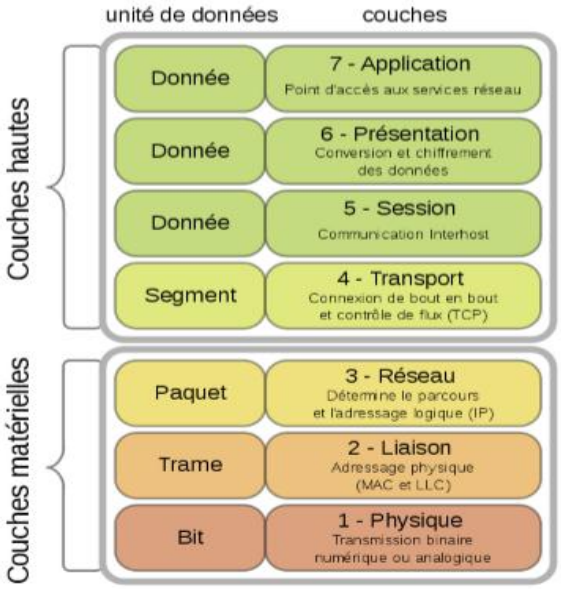


Figure 1.3 – Exemple d'attaque DDoS[20].

3.3Les types des attaques DDOS

Déférents types d'attaques DDoS ciblent déférents composants de la connexion réseau. Afin de comprendre le fonctionnement des déférentes attaques DDoS, il est nécessaire de comprendrecommentlesconnexionsréseausontétablies.Laconnexionréseausurinternet se compose de nombreux composants indépendants appelés « couches », chaque couche du modèle est conçue dans un but déférents. Le modèle OSI (illustrécidessous) est un cadre conceptuel pour décrire les connexions réseau dans sept couches déférentes :



Figur.1.4 – Le modèle OSI [25].

Les attaques DDoS sont classées dans trois grandes catégories en fonction de leur cible :

3.3.1. Attaques basées sur le volume

Il s'agit du type d'attaque DDoS le plus courant, les attaquant sessaient de consommer toute la bande passante du réseau en envoyant un nombre volumineux de paquets à un serveur victime Afin que cela devienne très difficile pour utilisateurs légitimes d'accéder au serveur en raison d'un blocage à routeurs d'interface. L'ampleur de ces types d'attaques est mesurée en bits par seconde. Exemple : Amplification DNS, UDP Flood, ICMP Flood, autres inondations de paquets usurpés [24].

- **L'attaque Amplification DNS** :

Il s'agit d'une attaque DDoS (déni de service distribué) volumétrique, fondée sur la réflexion, lors de laquelle un pirate exploite la fonctionnalité de résolveurs DNS ouverte pour surcharger un serveur ou un réseau cible avec une quantité de trafic amplifié, afin de rendre inaccessibles le serveur et son infrastructure environnante[25].

_ L'attaque par amplification DNS peut être divisée en quatre étapes :

1-Lepirateutiliseunpointdeterminaisoncompromispourenvoyerdespaquets UDP à un récurseur DNS à l'aide d'adresses IP usurpées. L'adresse usurpée contenue dans les paquets renvoie vers l'adresse IP réelle de la victime.

2- Chacun des paquets UDP envoie une requête à un résolveur DNS, en transmettant souvent un argument du type « ANY », afin de recevoir la réponse la plus volumineuse possible.

3- Après avoir reçu ces requêtes, le résolveur DNS (qui essaie d'exécutés son travail en répondant) envoie une réponse volumineuse à l'adresse IP usurpée

. 4- L'adresse IP de la cible reçoit la réponse et l'infrastructure de réseau environnante se retrouve submergée sous un flot de trafic, provoquant ainsi un déni de service [25].

Figure 1.5 – Exemple d'amplification[25].

- **Inondation UDP**

UDP Flood est un type d'attaque par déni de service dans laquelle un grand nombre de paquets UDP (User Datagramme Protocol) sont envoyés à un serveur ciblé dans le but d'écraser la capacité de cet appareil à traiter et à répondre.

Lepare-feuprotégeantleserveurciblépeutégalements'épuiseràlasuited'uneinondation UDP, entraînant un déni de service du trafic légitime [25].

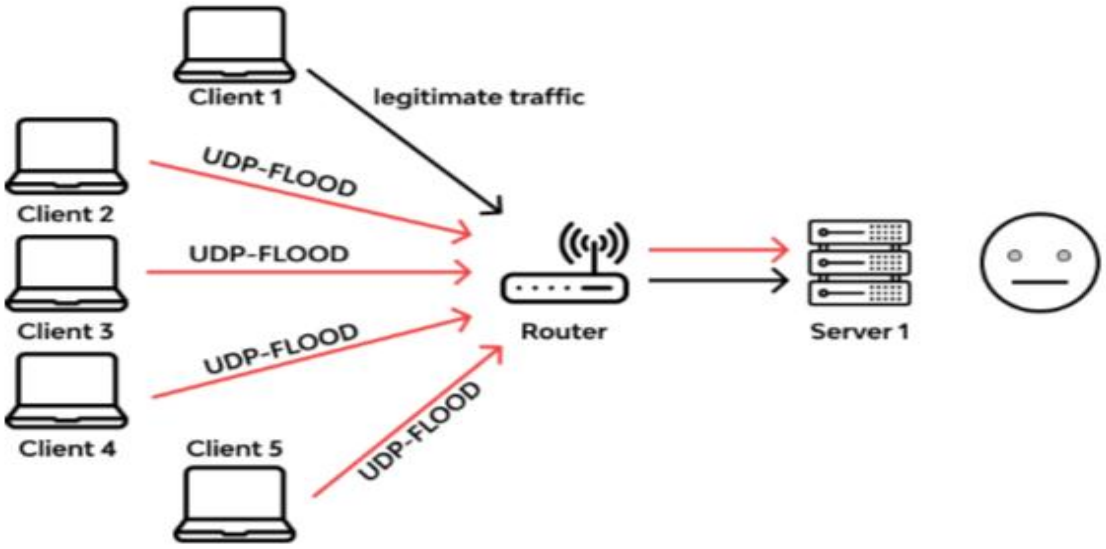


Figure .1.6 – Exemple d'attaques udp flood [26].

- **L'attaque Peer-To-Peer (P2P)**

De nos jours, les systèmes de partage de fichiers P2P ont été la cible d'attaques massivesquiexploitentdesboguesdanslesserveursP2PpourexécuterunDDoSattaque pour que l'attaquant P2P utilise des clients P2P connectés au P2P des hubs de partage de fichiers au lieu d'utiliser un bot pour attaquer l'hôte victime[26].

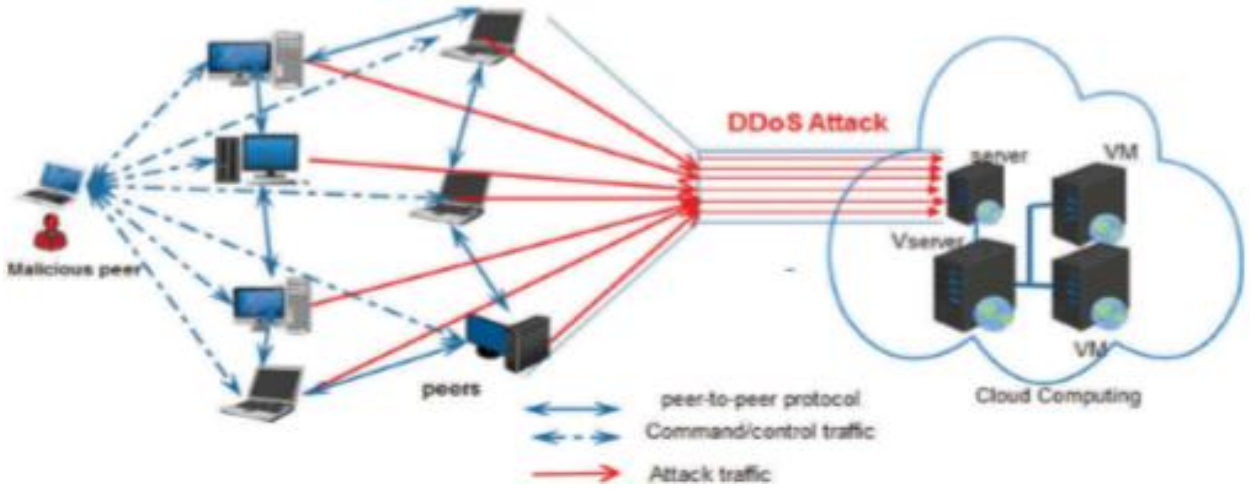


Figure 1.7 – L'architecture du mécanisme d'attaque P2P [26].

3.3.2 Attaques de protocole

Les attaques de protocole consomment les ressources de la victime machine comme la mémoire, la capacité de traitement, etc. Peut créer de longues files d'attente ou d'autres structures de données sur les périphériques réseau intermédiaires tels que les pare-feu, les routeurs ou équilibreurs de charge, etc. Ampleur du protocole les attaques sont mesurées en paquets par seconde. Exemple :SYN Flood, attaque de paquets fragmentés.

Exemple d'attaque de protocole : [25]

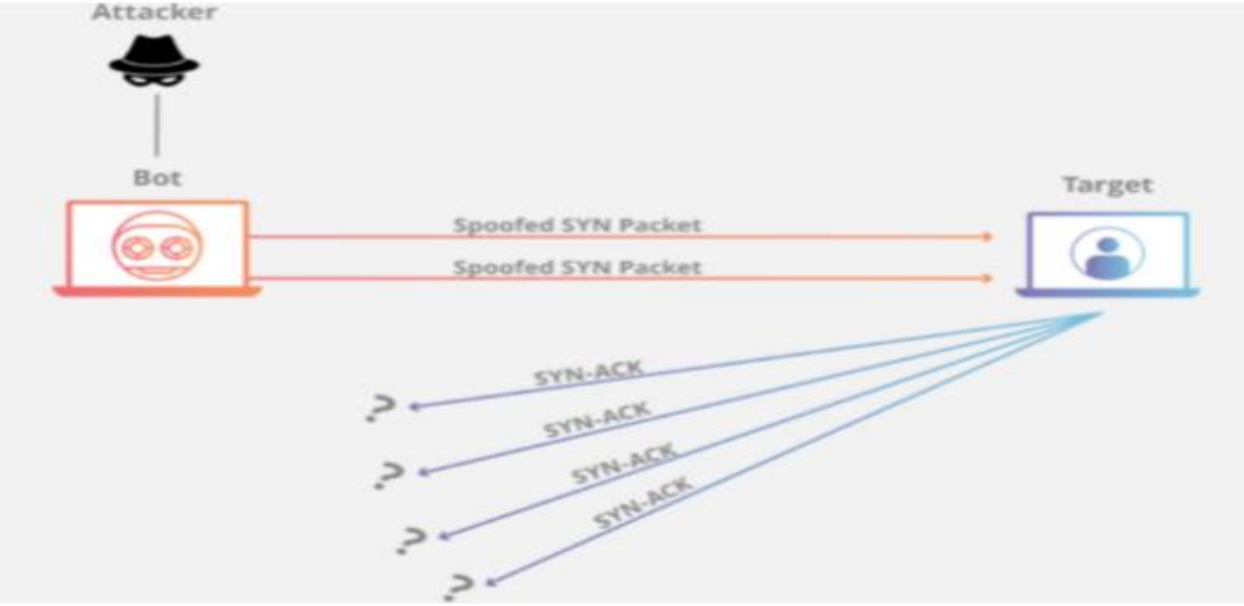


Figure 1.8 – Exemple d'attaque de protocole [25].

- **L'attaque SYN flood**

Une attaque SYN flood est un type d'attaque par déni de service (DDoS) qui vise à rendre un serveur indisponible pour le trafic légitime en consommant toutes les ressources serveur disponibles. En envoyant à plusieurs reprises des paquets de demande de connexion initiale (SYN), le pirate est en mesure de submerger tous les ports disponibles sur une machine serveur ciblée, ce qui oblige l'appareil ciblé à répondre

Lentement au trafic légitime, ou l'empêche totalement de répondre. l'attaques SYN flood fonctionne en utilisant le processus d'établisse mendé liaison des connexions TCP.

Généralement, uneconnexionTCPatroisprocessusindépendantspourétabliruneconnexion

. 1- Tout d'abord, le client envoie un paquet SYN au serveur pour établir une connexion.

2- Le serveur répond alors à ce paquet initial avec un paquet SYN/ACK pour confirmer la réception de la communication.

3- Enfin, le client renvoie un paquet ACK pour confirmer la réception du paquet depuis le serveur.

Après avoir terminé cette séquence d'envoi et de réception de paquets de données, la connexion TCP est ouverte et les données peuvent être envoyées et reçues [25].

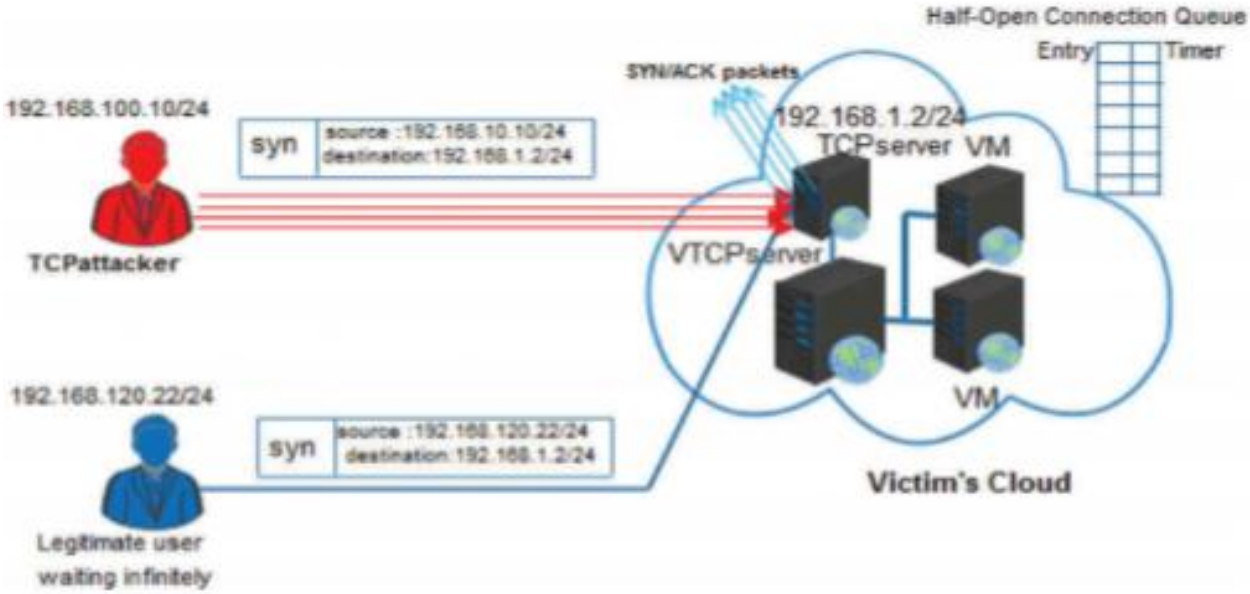


Figure 1.9 – Les détails du mécanisme d'attaque SYN Flood [26].

- **L'attaque Ping of death**

Une attaque Ping of death (PoD) est une attaque par déni de service (DoS), dans laquelle l'attaquant vise à perturber une machine ciblée en envoyant un paquet plus grand que la taille maximale autorisée, provoquant le blocage ou le crash de la machine cible. Lepingoriginaldel'attaquemortelleestmoinscourantaujourd'hui.Uneattaqueconnexe connue sous le nom d'attaque par inondation ICMP est plus répandue [25].

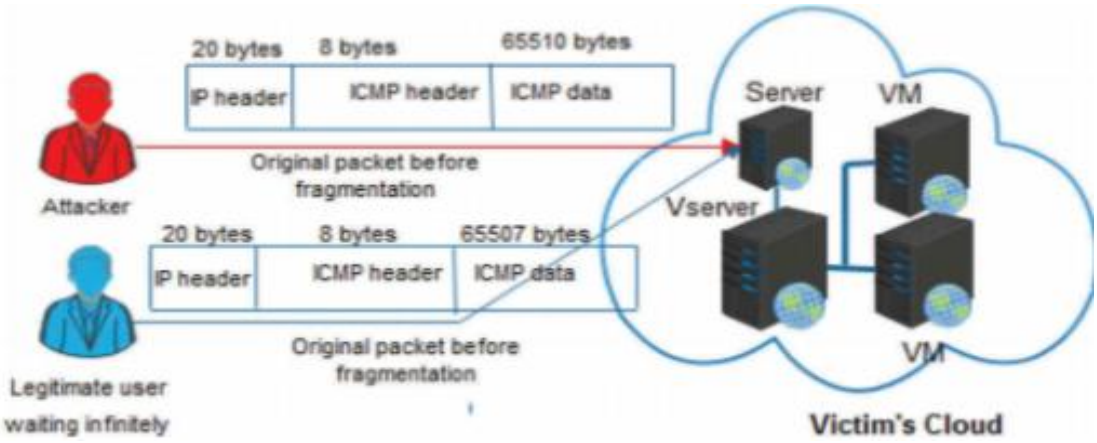


Figure 1.10 – L'architecture d'attaque Ping of death [26].

3.3.3 Attaques d'application

Considérées comme étant les types d'attaques DDoS les plus sérieuses et sophistiquées, ces attaques ciblent les applications web en exploitant les vulnérabilités qu'elles comportent.Aussiappelées«Attaquescouche7», les attaques d'application fonctionnent toujours de la même manière mais demandent beaucoup moins de force car elles ciblent les faiblesses au sein des serveurs ciblés.

Beaucoup moins de trafic web est nécessaire pour monopoliser les processus et protocoles sur ces points faibles, ce qui rend aussi l'attaque bien plus difficile à détecter en raison [24].

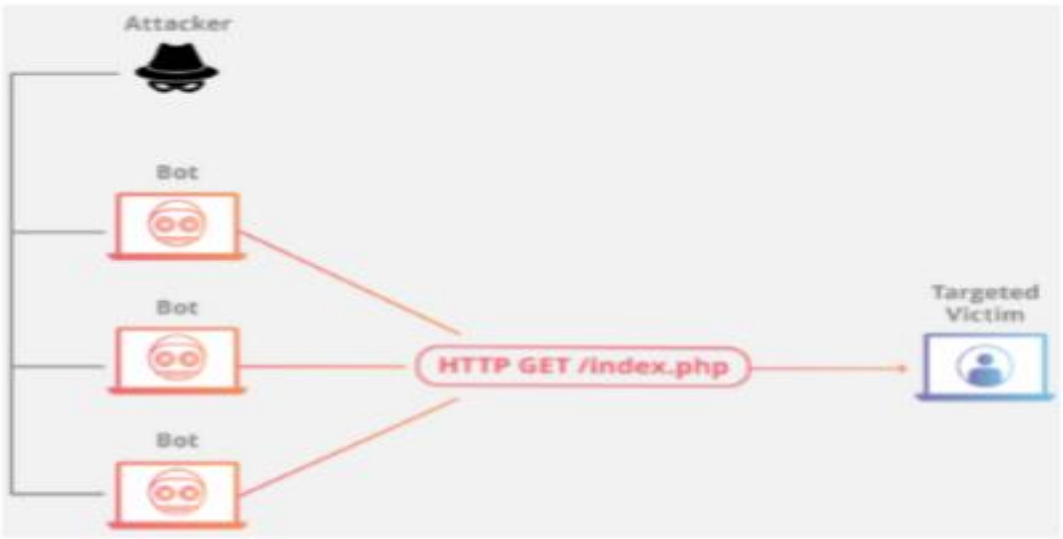


Figure 1.11 – Exemple d'attaque de la couche d'application [25].

- **L'attaque Back**

Pour lancer une attaque réussie sous cette forme, l'attaquant continu l'envoie d'un grand nombre de requêtes contenant de nombreuses barres obliques au serveur Web Apache. Chaque demande a besoin plus de temps de traitement qui ralentit offciellement le Web Apache serveur [26].

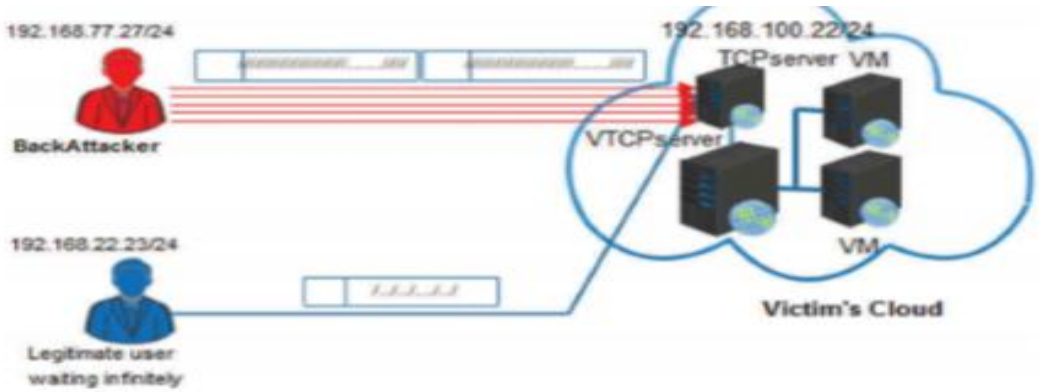


Figure 1.12 – l'architecture d'attaque retour [26].

- **L’attaque Slow loris**

Slow loris est une attaque de la couch applicative dans la quelle le cyber criminel envoie périodiquementdesrequêtesHTTPGETincomplètesmaislégitimesauserveurwebciblé.

Il maintient ainsi les connexions ouvertes et dévore lentement et méticuleusement les prises de connexion du serveur web, brouillant ainsi toutes les autres demandes légitimes [27].

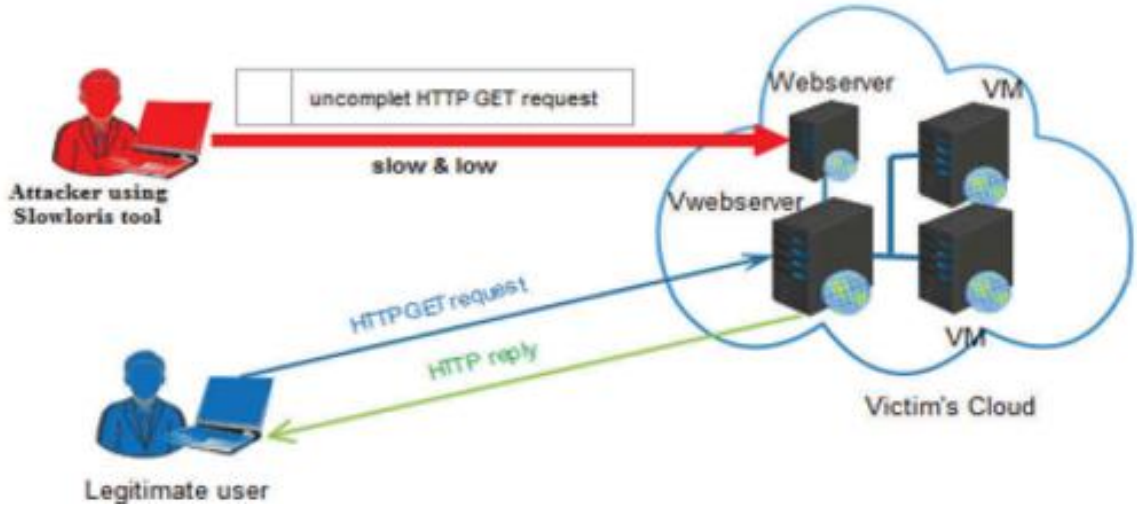


Figure1.13 – Les détails du mécanisme d'attaque de Slow loris [25].

4 Quelques vecteurs d’attaque

4.1 Les botnets

Les attaques DDoS peuvent être lancées à partir du réseau de la machine infectée Cela s’appelle un botnet. De nombreux outils disponibles en ligne peuvent tirer parti des botnets [28]. Dans ces dernières années, des services DDoS en ligne (communément appelés bootstrap ou stress) ont vu le jour [29]. Le prix de ces services permet utilisé par des individus et permet aux utilisateurs de lancer des attaques opposez-vous au but qu’il a choisi. En outre, certains programmes de démarrage fournissent également des services de test gratuit pendant quelques minutes. La diversité des outils et services permet lancer une attaque par déni de service distribué peut aider à augmenter beaucoup de ces attaques.

4.2 Les attaques basées sur la réflexion

Certaines attaques utilisent des ordinateurs accessibles sur internet et répondent demandes de n’importe quelle source : ce sont des réflecteurs. L’attaque de La réflexion consiste à envoyer des paquets à ces réflecteurs en utilisant les adresses IP suivantes : La victime est l’adresse IP source : c’est ce qu’on appelle l’usurpation d’adresse IP. Cette réaction de ces réflecteurs à la victime peut induire un trafic non sollicité de la destination de ce dernier [30].

Ce débit peut être suffisamment important pour saturer lien réseau de la victime, qui conduit à un déni de service. L'attaquant interroge le serveur en usurpant l'adresse IP ses victimes (203.0.113.2). Par conséquent, le serveur envoie la réponse à la demande faite par l'agresseur à la victime. Si le déni de service est effectif, Le trafic causé par ces réponses dépasse la bande passante du réseau disposition de la victime. Les attaques par réflexion impliquent généralement des protocoles basés sur des protocoles transmission UDP. En fait, le protocole UDP laisse la tâche à la couche application identifié la source afin que l'usurpation d'adresse IP puisse être effectuée. Au fait, UDP Il n'est pas nécessaire d'établir une session avant d'envoyer des données (contrairement à TCP). Cette fonctionnalité permet à l'attaquant d'utiliser UDP pour demander un service via un seul paquet de données et générer un la réponse du réflecteur. Les attaques par réflexion ne se limitent pas aux protocoles de transmission UDP. Par exemple, vous pouvez envoyer des paquets TCP SYN en usurpant l'adresse IP de la victime génère un paquet SYN-ACK en réponse à la cible. Enfin, il convient de noter que les attaques par réflexion peuvent être lancées à partir de botnets.

4.3 Les attaques basées sur l'amplification

Le but d'une attaque par lots est d'épuiser la bande passante réseau disponible pour rendre un ou plusieurs services inaccessibles.

Ce type d'attaque est généralement réalisé en utilisant les propriétés de certains protocoles pour maximiser le trafic généré. De plus, le but des attaques de volume est de produire très grand nombre de paquets de données par seconde pour saturer les ressources de traitement le but.

La réponse générée par certains protocoles est beaucoup plus grande que réclamer. Le nombre de paquets provoqués par la réponse peut également être plus grand dépassement du nombre de paquets requis pour envoyer la demande. L'amplification produit grâce à ces protocoles, il peut être utilisé pour mener des attaques de volume. Dans la plupart des cas, les attaques volumétriques utilisent la réflexion et l'amplification. De nombreux protocoles peuvent être utilisés pour implémenter ce genre d'attaque. Parmi eux, on peut notamment citer DNS (Domain Name System [30]), NTP (Network Time Protocol [31]), SNMP (Simple Network Management) Protocole [32]), SSDP (Simple Service Discovery Protocol) [33] ou CHARGEN (Protocole Générateur de Caractères [34]).

Il est important de noter qu'une entité peut être victime d'une attaque de volume utilisez le protocole, même s'il n'a aucun service actif exposé sur internet Basé sur le même accord.

4.4 Les attaques ciblant des applications

Certaines attaques visent à épuiser les capacités de traitement d'une cible. Par exemple, un attaquant peut chercher à atteindre la limite du nombre de connexions concurrentes qu'un

serveur web peut traiter. Dans ce cas, l'attaquant envoie en permanence un grand nombre de requêtes HTTP GET ou POST au serveur ciblé. Il est également possible d'envoyer des requêtes partielles, puis de transmettre la suite de ces requêtes à intervalles réguliers, dans le but de maintenir les connexions ouvertes le plus longtemps possible et d'éviter la fermeture des connexions au-delà d'un délai fixé [35].

D'autres types d'attaques applicatives cherchent à épuiser les ressources de calcul d'un serveur en initiant un grand nombre de sessions TLS, ou encore à tirer parti de faiblesses dans la conception d'une application web [36].

5 sécurités informatiques

Le cyber menace mondiale évolue à un rythme rapide, avec un nombre croissant de violations de données se produisant chaque année. Par conséquent, l'importance de la sécurité informatique a augmenté. La sécurité informatique exige la confidentialité, l'authentification, l'intégrité et la non-répudiation pour empêcher les pirates d'accéder aux données sans autorisation.

L'authentification est le processus de vérification de l'identité d'un utilisateur avant de donner accès au réseau. Il sert de première ligne de protection contre les pirates. La non-répudiation consiste à vérifier que les données ont été fournies avec les informations d'identification de l'utilisateur afin que les données puissent être identifiées avec l'expéditeur sans déni ni réputation. La confidentialité est la garantie que les données n'ont pu être consultées par personne d'autre que le destinataire prévu, garantissant que les données sont restées intactes jusqu'à leur réception . L'intégrité des données consiste à garantir que le contenu des données n'a pas été altéré pendant leur transport [36].

6 Techniques de sécurité

- Mettre constamment à jour votre système d'exploitation et vos applications.
- Utilisez un antivirus.
- Utilisez des mots de passe sécurisés (forts).
- N'ouvrez pas les pièces jointes d'expéditeurs inconnus Il n'est pas recommandé de cliquer sur des liens dans des e-mails provenant de sites Web inconnus ou d'expéditeurs inconnus.
- N'utilisez pas de réseaux Wifi non sécurisés dans des lieux publics.

7 Conclusion

Bien que les attaques DDoS soient en augmentation. Ce chapitre présente brièvement les attaques DDoS en informatique, suivi d'une classification des attaques et des différents types et contre-mesures pour atténuer les attaques DDoS. Cette enquête donne la technologie de détection, de prévention et de tolérance DDoS.

Chapitre 2 : SYSTÈMES DE DÉTECTION D'INTRUSION

1. Introduction

La détection d'intrusion est un élément crucial d'un système de protection et de sécurité. Parce que de nouvelles cyber attaques apparaissent chaque jour, les systèmes de détection d'intrusion (IDS) jouent un rôle principal dans la reconnaissance d'éventuelles cyber attaques sur un réseau ou un système et dans la fourniture de réponses appropriées. Les IDS doivent s'acclimater à ces nouvelles menaces (cyber attaques) et à leurs stratégies, tout en continuant à évoluer.

De nombreux outils sur le marché offrent désormais plusieurs niveaux de détection d'intrusion. Certaines solutions utilisent des signatures pour surveiller les attaques connues. Certaines plateformes assurent la surveillance des réseaux ; d'autres sont des systèmes basés sur l'hôte. Certaines solutions répondent à des avertissements spécifiques en fermant des services. Nous devons choisir avec soin les tactiques de détection d'intrusion pour garantir la sûreté et la sécurité des ressources de notre réseau contre les intrus indésirables.

Dans ce chapitre, nous allons présenter le Système de Détection d'Intrusion et son modèle général. Nous décrirons ensuite les types de systèmes de détection d'intrusion. Les techniques IDS sont abordées ci-dessous, avant d'introduire les techniques de détection des attaque DDOS et de conclure le chapitre

2 Définition

Les méthodes traditionnelles de protection contre les intrusions, telles que les pare-feu, la protection d'accès et le chiffrement, ont eu du mal à sécuriser complètement les réseaux et les serveurs contre les menaces et les logiciels malveillants de plus en plus avancés. En conséquence, les systèmes de détection d'intrusion (IDS) sont devenus un élément essentiel de l'infrastructure de défense pour détecter ces attaques avant qu'elles ne causent des dommages considérables [37].

Un système de détection d'intrusion (IDS) est un système logiciel ou matériel qui détecte l'accès illégal à un système informatique ou à un réseau [38].

3 Modèle général [39].

L'IDWG (Intrusion Détection Waring Group) de l'IETF a défini un modèle générique de détection d'intrusion qui représente la fonctionnalité commune à tous les IDS (qu'ils utilisent l'approche comportementale ou l'approche scénario).

La figure 2.1, proposée par l'IDWG (Intrusion Détection Working Group) montre le processus générique des systèmes de détection d'intrusion.

L'administrateur configure les différents composants (capteur(s), analyseur(s), gestionnaire(s)). Les capteurs accèdent aux données brutes, les filtrent et les formatent pour ne renvoyer que les événements intéressants à un analyseur. Les analyseurs utilisent ces événements pour décider si une intrusion est présente ou non et, si nécessaire, envoyer une alerte au gestionnaire (qui prévient l'opérateur humain). Une éventuelle réaction peut être effectuée automatiquement par le gestionnaire ou manuellement par l'opérateur.

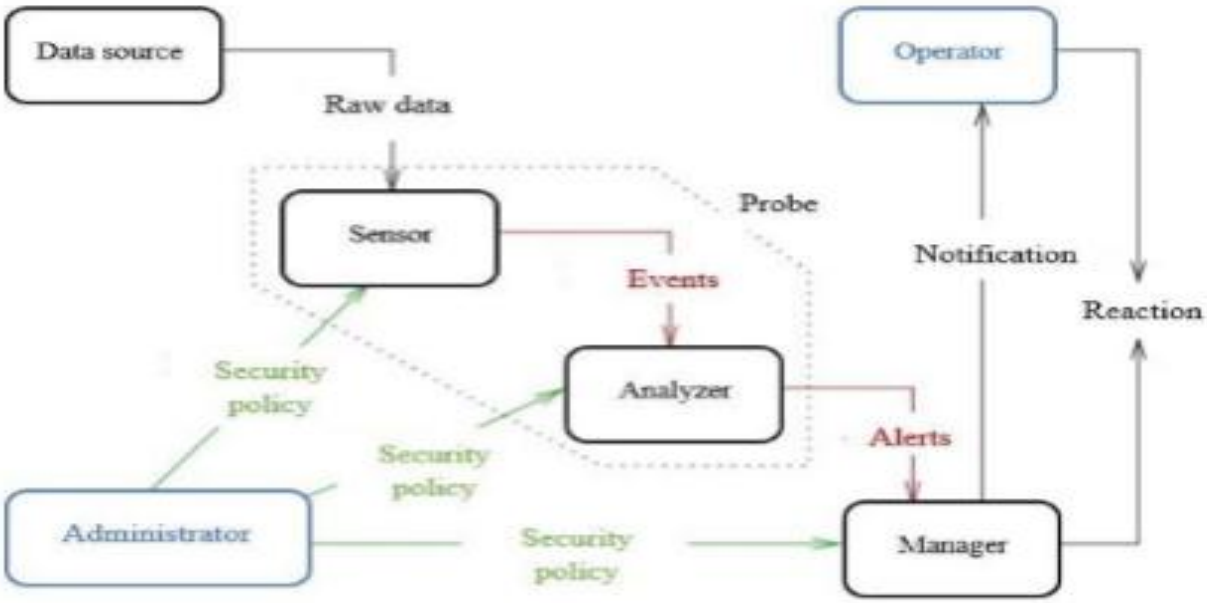


Figure 2.1 Modèle générique de détection d'intrusion proposé par l'IDWG [39].

4 types d'ID [40].

Un système de détection d'intrusion (IDS) détecte les intrusions indésirables dans les ordinateurs et les réseaux. Ces systèmes sont connus pour créer des alarmes (alertes) lorsqu'une intrusion est détectée. Les IDS sont classés en trois catégories, à savoir les systèmes de détection d'intrusion basés sur l'hôte (HIDS), les systèmes de détection d'intrusion basés sur le réseau (NIDS) et les IDS hybrides

4.1 IDS basé sur l'hôte (HIDS) [40].

HIDS est le premier type de détection d'intrusion qui a été développé. HIDS surveille et analyse les activités au niveau du système de l'hôte unique ou du système informatique interne, telles que la configuration du système.

HIDS est capable d'appliquer des politiques, d'analyser des journaux, de corréler des événements et de vérifier l'intégrité. Il a la capacité de déterminer si une attaque a réussi ou non.

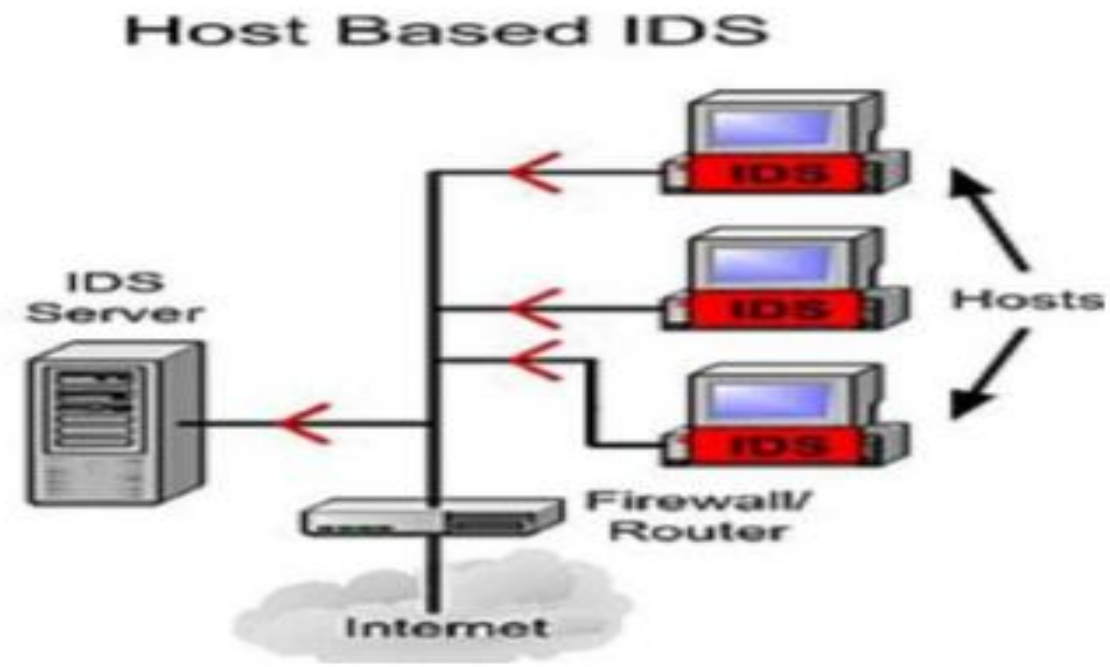


Figure 2. 2 IDS basé sur l'hôte [40].

4.2 IDS basé sur le réseau (NIDS) [40].

NIDS surveille et analyse les trafics réseau sur des segments de réseau particuliers pour détecter les activités suspectes.

La figure 2.3 illustre le NIDS et ses étapes de détection d'attaque. Les NIDS sont utilisés dans l'analyse au niveau des paquets pour l'ensemble des systèmes du segment de réseau en vérifiant le réseau de transport, l'IP, les en-têtes du paquet et les activités au niveau du protocole d'application pour détecter plusieurs attaques DOS basées sur IP.

Le NIDS peut détecter les attaques au fur et à mesure qu'elles se produisent. Cependant, NIDS n'indique pas si les attaques ont réussi ou non, car il n'analyse pas le système de journalisation.

Le problème du NIDS est qu'il a une visibilité limitée sur le système hôte et qu'il n'existe aucun mécanisme efficace pour analyser les données réseau cryptées pour détecter les attaques.

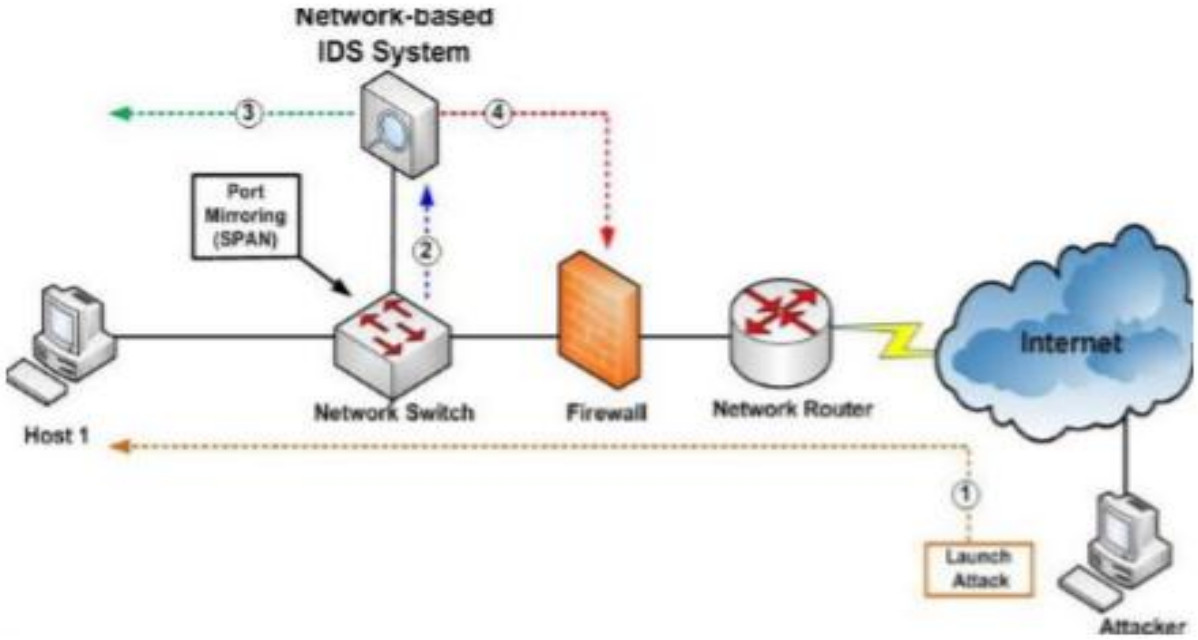


Figure 2. 3 IDS basé sur le réseau [40].

4.3 IDS hybrides

Hybride IDS combine deux ou plusieurs types d'IDS pour obtenir les avantages de l'IDS tout en effectuant une détection précise, comme Double Guardi qui utilise les ID d'hôte et l'ID de réseau. Cependant, il faut plus de temps pour analyser les données. La figure 2.4 illustre le système de détection d'intrusion hybride [40].

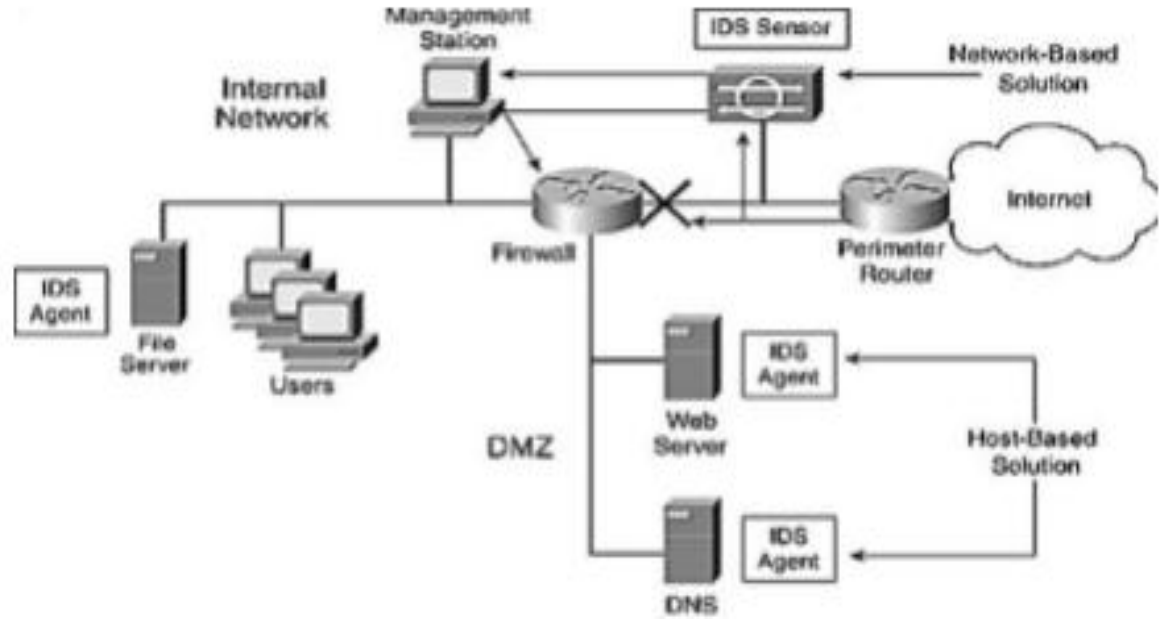


Figure 2. 4 IDS hybrides [40].

5 techniques IDS

De nombreuses techniques IDS sont utilisées dans la détection d'intrusions telles que la détection basée sur la signature, la détection d'anomalies, l'IDS basé sur le réseau de neurones artificiels (ANN), l'IDS basé sur la logique floue, l'IDS basé sur les règles d'association, l'IDS basé sur la machine à vecteurs de support (SVM), IDS basé sur un algorithme génétique (AG), techniques hybrides [41].

6 Conclusion

La sécurité du réseau est devenue un problème sérieux car plusieurs attaquants tentent d'attaquer les réseaux pour atteindre un objectif, comme l'économie. De nombreuses méthodes de protection du réseau ont été proposées, telles que les systèmes de détection d'intrusion, la cryptographie, les pare-feu, etc. Parmi ces outils de sécurité, la détection d'intrusion est généralement considérée comme l'une des méthodes les plus encourageantes pour nous protéger contre les cybers attaques nouvelles, dynamiques et complexes.

Dans ce chapitre, nous avons d'abord présenté les Systèmes de Détection d'Intrusion et leur modèle général. Nous avons ensuite discuté des trois types de systèmes de détection d'intrusion : Intrusion basée sur l'hôte.

Systèmes de détection d'intrusion basés sur le réseau (HIDS), systèmes de détection d'intrusion basés sur le réseau (NIDS) et systèmes hybrides de détection d'intrusion. Nous avons discuté des techniques des systèmes de détection d'intrusion avant d'introduire les techniques de détection des attaques DDoS. Dans le chapitre suivant, nous présenterons l'apprentissage automatique, l'apprentissage en profondeur et les réseaux de neurones.

Chapitre 3 : Machine Learning

1. Introduction

L'intelligence artificielle (IA) fait référence à la production de machines intelligentes comme l'esprit humain. L'intelligence artificielle concerne la conception et la mise en œuvre de systèmes de machines. Ces machines peuvent résoudre des problèmes qui nécessitent habituellement la capacité des humains.

Depuis l'article fondateur d'Alan Turing en 1950 sur la possibilité de programmer une machine électronique pour qu'elle agisse intelligemment, l'intelligence artificielle a connu au cours des dernières décennies une croissance rapide de la recherche et du développement.

Au XXIe siècle, l'intelligence artificielle (IA) est devenue un sujet de recherche essentiel dans tous les domaines tels que la science, le droit, la médecine, l'éducation, la comptabilité, les affaires, le marketing, la finance, l'économie et l'ingénierie.

L'apprentissage automatique est l'un des domaines du monde informatique actuel. De nombreuses recherches avaient été faites pour rendre les machines intelligentes. L'apprentissage est une activité humaine naturelle qui est également devenue un aspect crucial des ordinateurs. Les chercheurs ont déployé beaucoup d'efforts pour améliorer la précision des algorithmes d'apprentissage automatique.

L'apprentissage en profondeur (DL) devient de plus en plus crucial dans notre vie quotidienne. Il a déjà eu une influence significative dans plusieurs domaines, tels que les voitures autonomes et la reconnaissance vocale.

Dans ce chapitre, nous présenterons l'apprentissage automatique avant de discuter de la manière dont il passe à l'apprentissage en profondeur. Ensuite, nous présenterons l'apprentissage en profondeur avant de terminer le chapitre sur les réseaux de neurones dans la dernière section.

2 Machine Learning

2.1 Définition [42]

Le Machine Learning, aussi appelé apprentissage automatique en français, est un sous-domaine de l'intelligence artificielle. Il s'agit d'étudier les algorithmes informatiques qui permettent aux systèmes d'apprendre et de s'améliorer automatiquement à partir de chaque expérience.

Selon l'approche d'apprentissage, le type de données entrées et sorties et le type de problème résolu, les algorithmes d'apprentissage automatique ont été divisés en plusieurs types : apprentissage supervisé, non supervisé, semi-supervisé et apprentissage renforcé.

2.2 Types Machine Learning (d'apprentissage automatique)

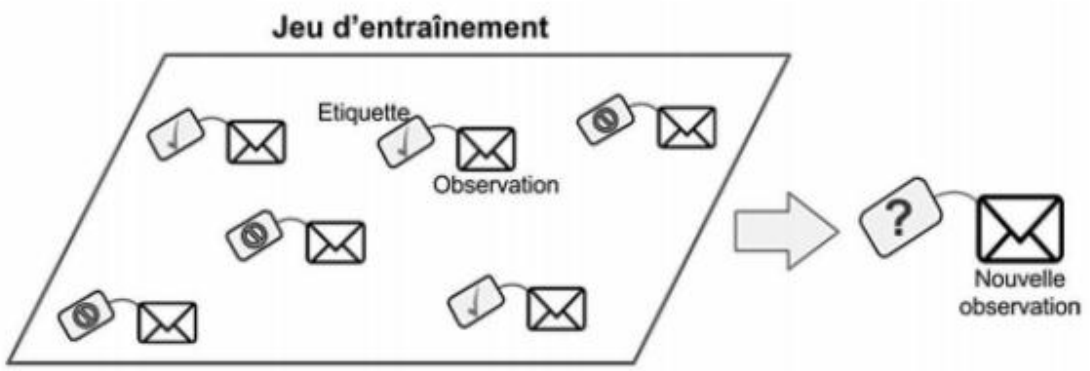
2.2.1 Apprentissage supervisé

L'apprentissage supervisé est une catégorie d'apprentissage automatique dans laquelle les algorithmes apprennent la variable d'entrée (X) en tant que superviseur ou enseignant pour prédire la variable de sortie (Y). Cette catégorie est appelée apprentissage supervisé car l'algorithme peut apprendre de l'ensemble de données d'entraînement étiqueté, tandis que arrêter lorsque l'algorithme at. teint un processus d'apprentissage niveau de performance approprié. Le plus couramment.

Utilisé les algorithmes d'apprentissage supervisé incluent des vecteurs de support Machine (SVM), régression linéaire, régression logistique, naïf Algorithme bayé sien et k voisin le plus proche (KNN).

• Un exemple d'utilisation de l'apprentissage supervisé est le filtre anti-spam, l'apprentissage s'effectue à l'aide de nombreux exemples d'é-mails qu'on a étiqueté spam ou normal, a partir de cela, le filtre doit alors être capable de classer de nouveaux e-mails.

• Un autre exemple consiste à prédire le prix d'une voiture à partir des valeurs d'un certain nombre d'attributs ou variables qu'on appelle caractéristiques d'une observation ou features en anglais. Ces variables peuvent être le kilométrage, l'âge, la marque, etc. Ils sont également appelés variables explicatives ou prédictives. L'entraînement se fait alors à partir de ces variables et des étiquettes. La catégorie de la variable prédite Y fait décliner. [43] [44] [45].



La figure 3.1 – Un jeu d'entraînement étiqueté pour un apprentissage supervisé [45].

Existe deux types d'apprentissage supervisé :

- **Classification** :

Un problème de classification survient lorsque la variable de sortie est une catégorie, telle que «rouge», «bleu» ou «maladie» et «pas de maladie». Exemples :

• En finance et dans le secteur bancaire pour la détection de la fraude par carte de crédit (fraude, pas fraude).

• Détection de courrier électronique indésirable (spam, pas spam).

• Dans le domaine du marketing utilisé pour l'analyse du sentiment de texte (heureux, pas heureux).

• En médecine, pour prédire si un patient a une maladie particulière ou non.

- **Régression** : Un problème de régression se pose lorsque la variable de sortie est une Valeur réelle, telle que «dollars» ou «poids».Exemples [46]

• Prédire le prix de l'immobilier.

• Prédire le cours de bourse

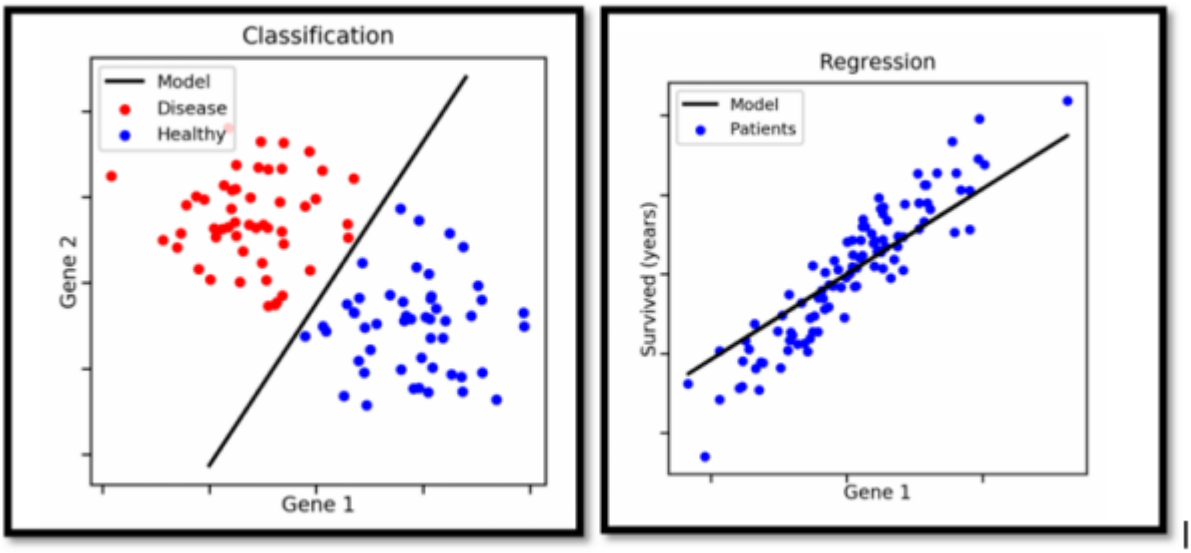


Figure 3.2 – La Classification et la Régression [46].

Problème Linéaire et Non-Linéaire :

Les méthodes de classification supervisée peuvent être basées sur :

- des hypothèses probabilistes (cas du classifieur naïf bayésien).

- des notions de proximité (exemple, k plus proches voisins).

- des recherches dans des espaces d'hypothèses (exemple, arbres de décisions). Selon le problème, il faut pouvoir choisir un classificateur approprié, c'est-à-dire Il sera possible de séparer au maximum les données d'entraînement. Si les exemples de classes différentes sont, on dit qu'un problème est linéairement séparable Peut être complètement séparé par un hyperplan (appelé hyperplan de séparation ou séparateur), Ce type de problème est résolu par un classificateur assez simple dont le but est de trouver Équation de l'hyperplan séparant. Mais, le

problème peut également être non séparable de manière linéaire comme illustré dans la figure 3.3. Dans ce cas, il faut utiliser d'autres types de classifieurs, souvent plus longs à paramétrer, mais qui obtiennent des résultats plus précis. [47]

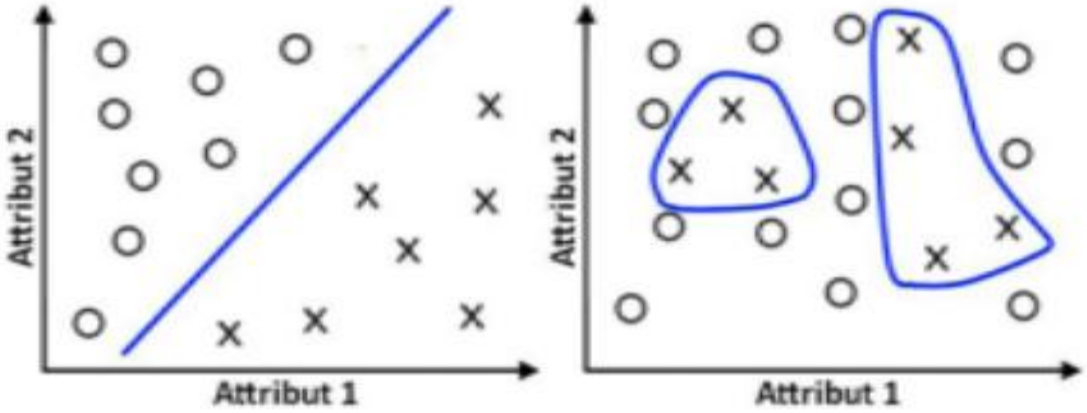
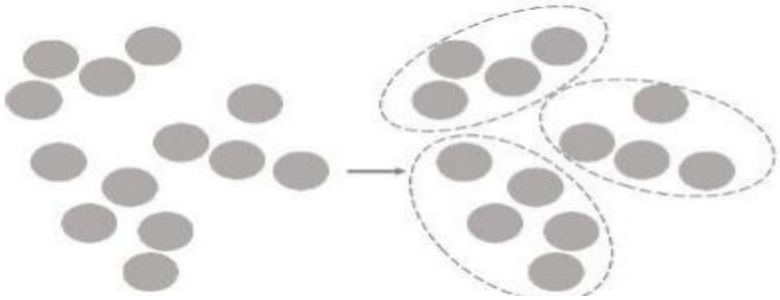


Figure 3.3 – A Gauche : Problème linéairement séparable (Frontière linéaire). A Droite : Problème non linéairement séparable [47].

2.2.2 Apprentissage non supervisé

L'apprentissage non supervisé est utilisé lorsque les données ne sont utilisées qu'en entrée et qu'il n'y a pas de variable de sortie correspondant à ces données. Pour en savoir plus sur les caractéristiques des données, un tel algorithme modélise les modèles sous-jacents de ces données.

L'un des algorithmes non supervisés les plus courants est le clustering. Cette technique découvre des groupes inhérents dans les données, puis les utilise pour la prédiction de la sortie pour les entrées invisibles. La figure 3.4 montre un exemple d'apprentissage non supervisé [42].



La figure 3.4 Présentation de l'apprentissage non supervisé. Les échantillons d'entrée sont regroupés en grappes en fonction des modèles sous-jacents

2.2.3 Apprentissage semi-supervisé

Ce type d'algorithme est un intermédiaire entre les techniques d'apprentissage non supervisé et supervisé. Cet algorithme s'entraîne en utilisant une combinaison de données non étiquetées

(petite quantité) et étiquetées (petite quantité). L'algorithme d'apprentissage non supervisé regroupe les premières données similaires, puis il étiquette les données non étiquetées au repos en utilisant les données étiquetées existantes. La figure 3. 5 montre un exemple d'apprentissage semi-supervisé. [42]

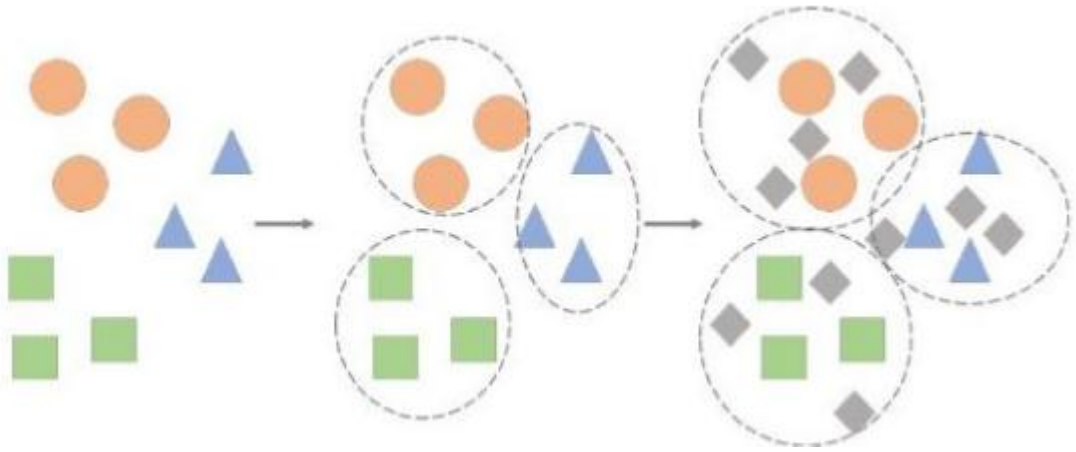


Figure 3.5 Aperçu de l'apprentissage semi-supervisé. Les clusters formés par une grande quantité de données non étiquetées sont utilisés pour classer une quantité limitée de données étiquetées [42].

2.2.4 Apprentissage par renforcement

L'apprentissage par renforcement est utilisé lorsque l'objectif est de prendre une série de décisions qui conduisent à une récompense finale. Un agent artificiel reçoit soit des pénalités, soit des récompenses en fonction des actions qu'il accomplit au cours du processus d'apprentissage. L'apprentissage par renforcement vise à maximiser la récompense totale. La figure 3. 6 montre un exemple d'apprentissage par renforcement.

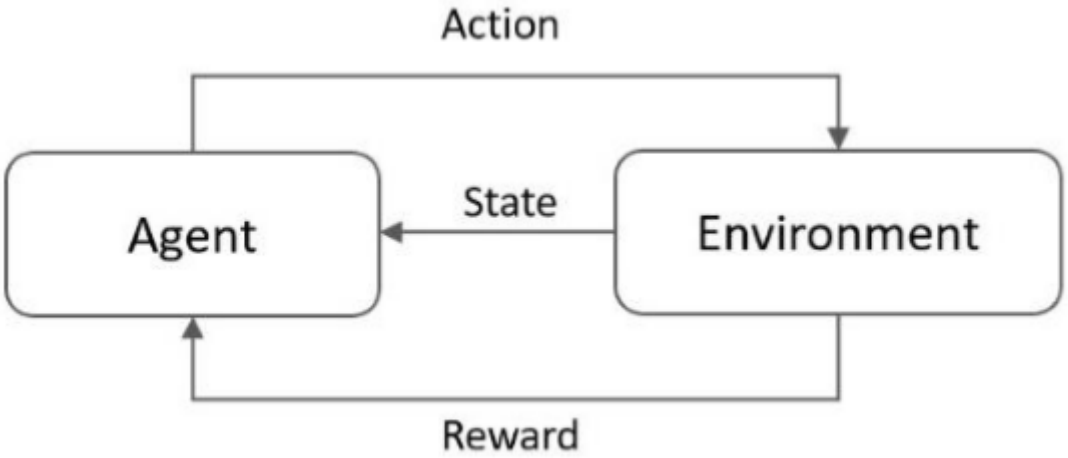


Figure 3.6 Vue d'ensemble de l'apprentissage par renforcement. Un agent observe l'état de l'environnement et effectue des actions pour maximiser une récompense globale [42].

3 Algorithmes de Classification :

La Classification est une méthode d'exploration de données utilisée pour attribuer des instances de données à l'une des quelques catégories. De nombreux algorithmes de Classification ont été développés pour se surpasser. Ils utilisent toues des techniques mathématiques pour fonctionner, telles que les arbres de décision, la programmation linéaire et les réseaux de neurones. Ces techniques analys ent les données disponibles de diverses manières pour faire des prédictions.Cependant, certains sont plus couramment utilisées que d'autres :

3.1 Naïve bayes:

Naïve Bayes Classifier est un algorithme populaire en Machine Learning. C'est un algorithme du Supervisé Learning utilisé pour la Classification. Il est particulièrement utile pour les problématiques de Classification de texte. Un exemple d'utilisation du NaiveBayes est celui du filtre anti-spam [48].

Byes theorem: P (H/X) = P(X/H.P (H) / P (H)

X : Données avec des classes qui n'ont pas connu

H : L'hypothèse des données est une classe spécifique

P(H|X) : La probabilité de l'hypothèse H est basée sur la condition X (probabilité a priori)

P (H) : Probabilité de l'hypothèse H (probabilité a priori)

P (X|H) : Probabilité X basée sur la condition de l'hypothèse H (la probabilité que l'évenement X se réalise sachant que l'évenement H s'est déjà réalisé).

P (X) : Probabilité X. Où C et X sont deux événements (par exemple, la probabilité que le train arrivera à l'heure étant donné que le temps est pluvieux). De tels classificateur naïfs de Bayes utilisent la probabilité théorie pour trouver la Classification la plus probable d'une instance (non classée). L'algorithme fectue positivement avec des données catégorielles mais malsain a données numériques dans l'ensemble d'apprentissage (the training set).

3.2 L'algorithme K Nearst Neighbors (K-NN) :

L'algorithme des K plus proches voisins ou K-nearest Neighbors (kNN) est un algorithme de Machine Learning qui appartient à la classe des algorithmes d'apprentissage

Supervisé simple et facile à mettre en œuvre qui peut être utilisé pour résoudre les problèmes de classification et de régression [49]. Pour exécutés une prédiction, l'algorithme K-NN va se baser sur le jeu de données en entier. En émets, pour une observation, qui ne fait pas parti du jeu de données, qu'on souhaite prédire [50].

L’algorithme va chercher les K instances du jeu de données les plus proches de notre observation. Ensuite pour ces K voisins, l’algorithme se basera sur leurs variables de sortie (output variable) y pour calculer la valeur de la variable y de l’observation qu’on souhaite prédire. En général, pour définir la distance entre deux objets x et y, la formule de distance euclidienne est utilisée dans l’équation suivante :

D(x, y)=v u u t n X j=1 (xj −yj) 2 par ailleurs :

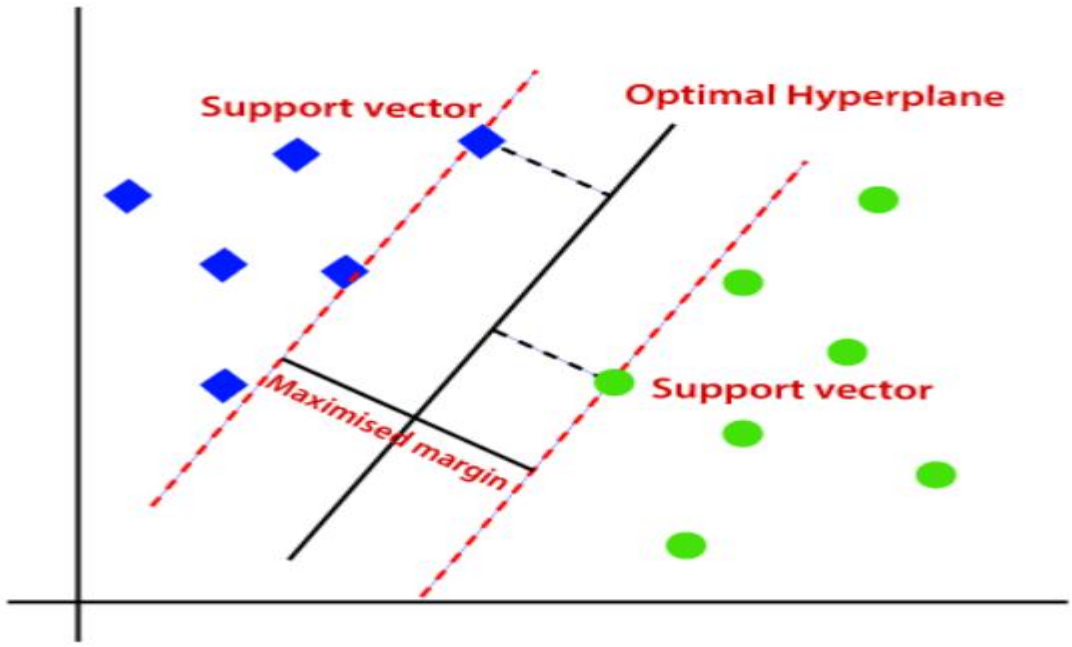
• Si K-NN est utilisé pour la régression, c’est la moyenne (oulamédiane) des variables y des K plus proches observations qui servira pour la prédiction.

• Si K-NN est utilisé pour la classification, c’est le mode des variables y des K plus proches observations qui servira pour la prédiction.

3.3 Support vecteur Machine(SVM)

3.3.1 Définition

Les machines à vecteurs de support, ou support vector machine (SVM), sont des modèles de machine Learning supervisés centrés sur la résolution de problèmes de discrimination et de régression mathématiques. Elles ont été conceptualisées dans les années 1990 à partir d'une théorie d'apprentissage statistique développée par les informaticiens russes Vladimir Vapnik et Alexey Chevronnais : la théorie de Vapnik-Chevronnais. Ce modèle a été rapidement adopté en raison de sa capacité à travailler avec des données de grandes dimensions, ses garanties théoriques et les bons résultats réalisés en pratique. Requérant un faible nombre de paramètres, les SVM sont appréciées pour leur simplicité d’usage [51].



Principe du Séparateur à Vaste Marge (SVM) [51].

3.3.2 Comment fonctionnent les SVM ?

Le principe des SVM consiste à ramener un problème de classification ou de discrimination à un hyperplan (features space) dans lequel les données sont séparées en plusieurs classes dont la frontière est la plus éloignée possible des points de données (ou "marge maximale"). D'où l'autre nom attribué aux SVM : les séparateurs à vaste marge. Le concept de frontière implique que les données soient linéairement séparables. Pour y parvenir, les support vector machines font appel à des noyaux, c'est-à-dire des fonctions mathématiques permettant de projeter et séparer les données dans l'espace vectoriel, les "vecteurs de support" étant les données les plus proches de la frontière. C'est la frontière la plus éloignée de tous les points d'entraînement qui est optimale, et qui présente donc la meilleure capacité de généralisation [52].

3.3.3 Une SVM peut-elle permettre de résoudre une régression ?

Oui. En 1996, Vladimir Vapnik, en collaboration avec d'autres mathématiciens (Harris Drucker, Chris Burges, Linda Kaufman et Alex Smola), a proposé une méthode permettant d'utiliser des SVM pour résoudre des problèmes de régression. Cela est possible là-encore en ayant recours à l'astuce du noyau, très utilisée en machine Learning car elle permet d'utiliser des classifieurs linéaires et de s'en servir pour résoudre un problème non linéaire. La version régression des SVM est appelée en anglais : Support Vector Régression (SVR) [52].

3.3.4 Comment fonctionnent les SVM ?

Le principe des SVM consiste à ramener un problème de classification ou de discrimination à un hyperplan (features space) dans lequel les données sont séparées en plusieurs classes dont la frontière est la plus éloignée possible des points de données (ou "marge maximale"). D'où l'autre nom attribué aux SVM : les séparateurs à vaste marge. Le concept de frontière implique que les données soient linéairement séparables. Pour y parvenir, les supports Vector machines font appel à des noyaux, c'est-à-dire des fonctions mathématiques permettant de projeter et séparer les données dans l'espace vectoriel, les "vecteurs de support" étant les données les plus proches de la frontière. C'est la frontière la plus éloignée de tous les points d'entraînement qui est optimale, et qui présente donc la meilleure capacité de généralisation [52].

3.4 Random Forest Classifier :

Random Forest Classifier est un algorithme d'apprentissage supervisé et composé de plusieurs arbres de décision. En faisant la moyenne de l'impact de plusieurs arbres de décision, les forêts aléatoires tendent à améliorer la prédiction [53].

Il fonctionne en quatre étapes :

1-Sélectionnez des échantillons aléatoires à partir d'un ensemble de données donné.

2- Construire un arbre de décision pour chaque échantillon et obtenez un résultat de prédiction de chaque arbre de décision.

3- Exécutés un vote pour chaque résultat prévu.

4- Sélectionnez le résultat de prédiction avec le plus votes comme prédiction finale.

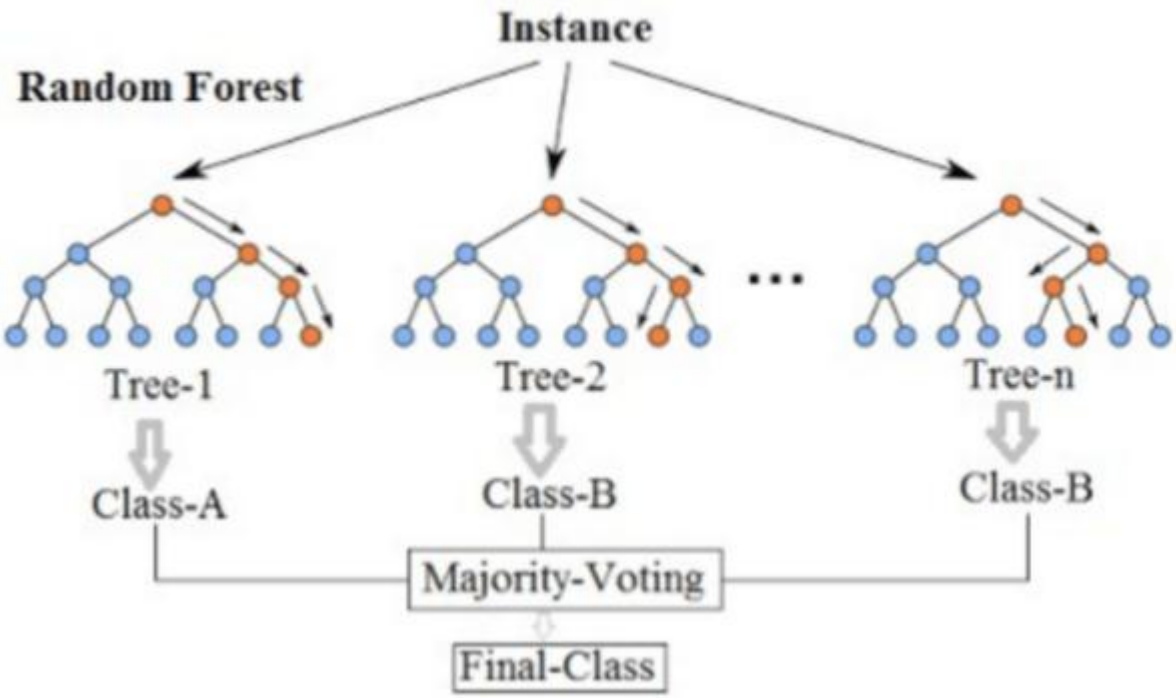


Figure 3.7 – Random Forest Classifier [53].

4 Deep Learning (Apprentissage en profondeur)

4.1 Définition

Le *deep Learning* ou apprentissage profond est un type d'intelligence artificielle dérivé du *machine Learning* (apprentissage automatique) où la machine est capable d'apprendre par elle-même, contrairement à la programmation où elle se contente d'exécuter à la lettre des règles prédéterminées [55].

4.2 Fonctionnement du deep Learning

Le deep Learning s'appuie sur un réseau de neurones artificiels s'inspirant du cerveau humain. Ce réseau est composé de dizaines voire de centaines de « couches » de neurones, chacune recevant et interprétant les informations de la couche précédente. Le système apprendra par exemple à reconnaître les lettres avant de s'attaquer aux mots dans un texte, ou détermine s'il y a un visage sur une photo avant de découvrir de quelle personne il s'agit À chaque étape, les « mauvaises »

réponses sont éliminées et renvoyées vers les niveaux en amont pour ajuster le modèle mathématique. Au fur et à mesure, le programme réorganise les informations en blocs plus complexes. Lorsque ce modèle est par la suite appliqué à d'autres cas, il est normalement capable de reconnaître un chat sans que personne ne lui ait jamais indiqué qu'il n'ai jamais appris le concept de chat. Les données de départ sont essentielles : plus le système accumule d'expériences différentes, plus il sera performant.

4.3 Applications du deep Learning [54]

Le deep Learning est utilisé dans de nombreux domaines :

- Reconnaissance d'image,
- Traduction automatique,
- Voiture autonome,
- Diagnostic médical,
- Recommandations personnalisées,
- Modération automatique des réseaux sociaux,
- Prédiction financière et trading automatisé,
- Identification de pièces défectueuses,
- Détection de malwares ou de fraudes,
- Chabots (agents conversationnels),
- Exploration spatiale,
- Robots intelligents.

C'est aussi grâce au deep Learning que l'intelligence artificielle de Google Alpha Go a réussi à battre les meilleurs champions de Go en 2016. Le moteur de recherche du géant américain est lui-même de plus en plus basé sur l'apprentissage par deep Learning plutôt que sur des règles .

5 Métriques utilisés :

5.1 Confusion matrix [57] :

Une Confusion matrix (matrice de confusion) ou tableau de contingence est un outil permettant de mesurer les performances d'un modèle de Machine Learning en vérifiant notamment à quelle fréquence ses prédictions sont exactes par rapport à la réalité dans des problèmes de classification. Ainsi, dans la classification binaire, le nombre de vrai négatif sont C1, 0, le vrai positif C1, 1, et les faux positifs C0, 1.

5.2 Accuracy :

Dans la classification multi-étiquette, cette fonction calcule la précision des sous-ensembles: l'ensemble d'étiquettes prédites pour un échantillon doit correspondre exactement a l'ensemble d'étiquettes correspondant dans y-true.

$$Accuracy = \frac{vraiPositif + vraiNégatif}{total}$$

5.3 Classification report [56] **:**

esklearn.metrics.classification-report : résumé textuel de la précision, rappel, score f1, pour chaque classe. Dictionnaire retourné si output-dict. est true. Le dictionnaire a la structure suivante :

• **précision** :

La précision est le rapport entre les vrais positifs et le total des vrais positifs et des faux positifs. La précision regarde pour voir combien de Junk positifs ont été jetés dans le mélange. S'il n'y a pas de mauvais positifs (ces fps), alors le modèle a eu la précision 100%. Plus il y a de fps qui entrent dans le mélange, plus la précision sera laide [56] . Mathématiquement

$$précision = \frac{tp}{tp + fp}$$

• **Recall :**

Le rappel est la mesure de notre modèle identifiant correctement les vrais positifs. Ainsi, pour tous les patients qui soufrent réellement d'une maladie cardiaque, le rappel nous indique combien nous avons correctement identités comme ayant une maladie cardiaque [56]. Mathématiquement

$$Recall = \frac{tp}{tp + fn}$$

• **F1 Score :**

Le score F1 est la moyenne harmonique de la précision et du rappel'Recall'

$$F1Score = \frac{2(p * r)}{p + r}$$

6 Conclusion

Ce chapitre nous a permis d'avoir une idée globale sur l'intelligence artificielle, l'importance de machine Learning, le types de systèmes d'apprentissage et les algorithmes de classification avec

leurs avantages et inconvénients. Le prochain chapitre va présenter la conception et la réalisation de notre projet.

Chapitre 4 : Réalisation

1 Introduction

Comme nous l'avons dit plus tôt dans le chapitre 3, l'apprentissage automatique est partout, donc dans ce chapitre, nous avons par couru les étapes d'un projet d'apprentissage automatique dans le but de détecter les attaques DDoS. Et obtenir de bons résultats pour cette découverte.

2 Conception de notre solution :

Comme notre solution propos de la détection des attaques DDOS basé sur le machine Learning, nous devons tout d'abord passer par les étapes de machine Learning (nettoyage, prétraitement des données, normalisation, apprentissage, validation et exécution), le fonctionnement du système de détection d'attaque proposé est illustré dans un organigramme présenté dans la figure suivante :

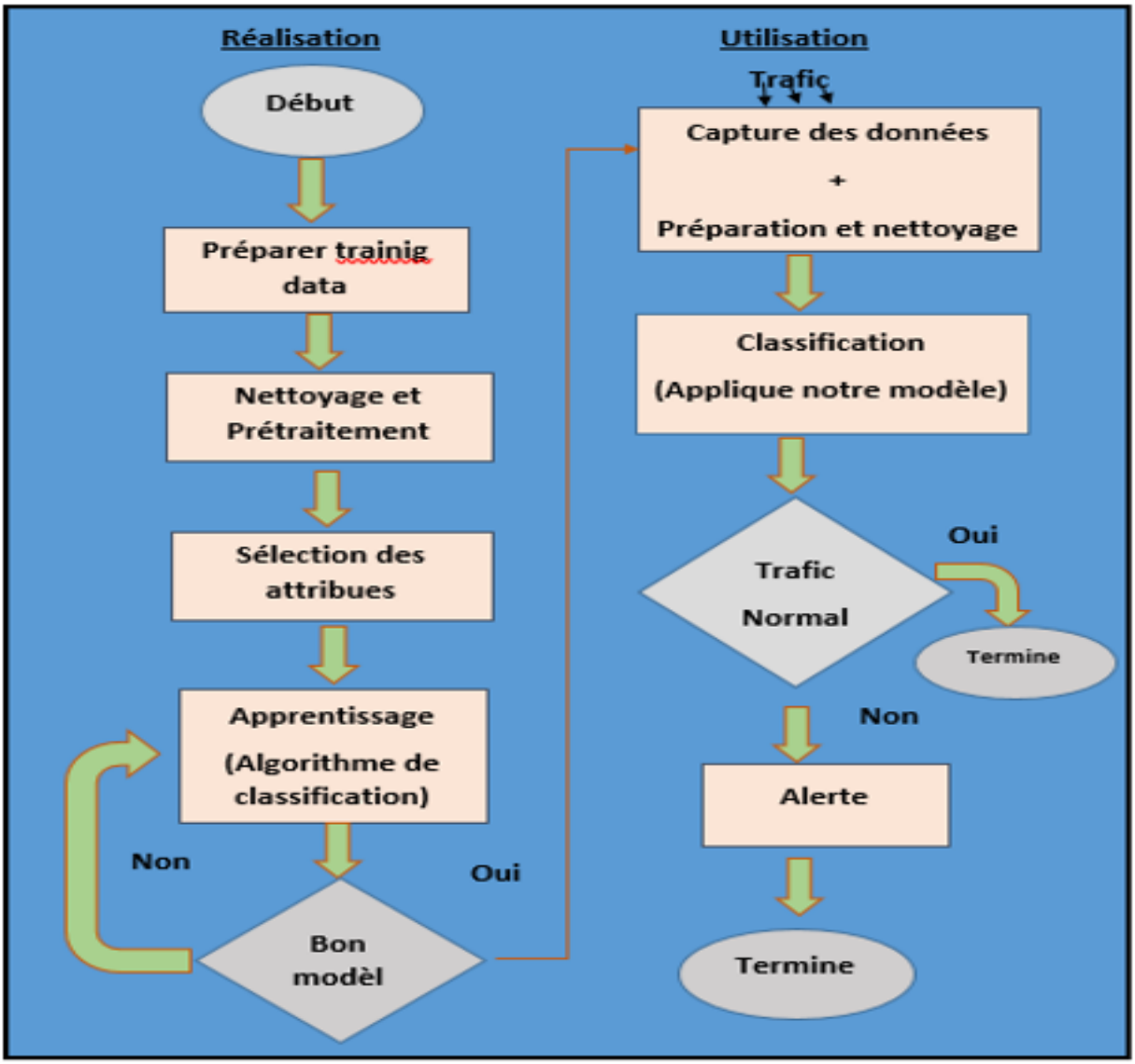


Figure 4.1 – Les étapes de machine Learning [58].

3 Choix du data set :

Cet ensemble de données a été créé par l'Université Bennett. Il s'agit d'un ensemble de données spécifique au SDN généré à l'aide d'un émulateur mini net et utilisé pour la classification du trafic par des algorithmes d'apprentissage automatique et d'apprentissage en profondeur. Le projet commence par créer dix topologies en mini net dans lesquelles les commutateurs sont connectés à un seul contrôleur Rye. La simulation de réseau s'exécute pour le trafic bénin TCP, UDP et ICMP et le trafic malveillant qui est la collection d'attaques TCP Syn., d'attaques UDP Flood, d'attaques ICMP. Au total, 21 fonctionnalités sont disponibles dans l'ensemble de données dans lequel certaines sont extraites des commutateurs et d'autres sont calculées [58]

Les attributs de notre Base de données [58] : Le tableau suivant donne la description des attributs de fi data SDN avec l'explication de chaque attribut

dt	affiche la date et l'heure
SRC	Adresse IP source
DST	Adresse IP destinations
pk count	Nombre des paquets
bytecount	Nombre des bytes
Dur_sec	La durée en second
Dur_nsec	la durée en nanosecondes
Tot_dur	durée totale est la somme de duration_sec et durstaion_nsec
flow	débit de paquets de flux
Packets	le nombre de paquets
pktperflow	le nombre de paquets pendant un seul flux
byteperflow	le nombre des bytes pendant un seul flux
pktrate	moyen du nombre de paquets
Port_no	le numéro de port
Tx_bytes	le nombre d'octets transférés depuis le commutateur
Rx_bytes	le nombre d'octets reçus sur le port du commutateur
Tx_kbps	le nombre kilooctets transférés depuis le commutateur
Rx_kbps	le nombre kilooctets reçus sur le port du commutateur

Tot_kbps	le nombre kilooctets reçus sur le port du commutateur
label	0 ou 1 c'est-a-dire BENIGN Ou DDoS
Protocol	Numéro du Protocol utilisé

Le tableau 4.1 : la description de chaque attribut.

4 Implémentions :

Voici les étapes que nous avons suivie pour faire l'implémentions de notre projet :

4.1 Prétraitement :

Le prétraitement c'est l'étape qui consiste a préparé nos données avant de les fournir à la machine pour son apprentissage, notre objectif et de :

1. Mètre les données dans un format propice au machine Learning :

• Préparation du data set : Nous avons formé la base de données accumulant 5% de chaque fichiers puis nous l'avons enregistré dans un fichier csv. Cette 'data set' l'ensemble du training est : 72687 et l'ensemble du test est : 31152.

- Après avoir éliminé les valeurs ou le contraste égal à zéro, vous obtiendrez 21 variables.

```
print(X_train.shape)
print('X test: ', X_test.shape)
print('Y_train: ', Y_train.shape)
```

```
(72687, 21)
X test:  (31152, 21)
Y_train:  (72687,)
```

Figure 4.2 – Le nombres d'attributs.

- Pour former et évaluer notre modèle, nous avons divisé au hasard l'ensemble de données à l'aide de la fonction «Train_test_split» en deux parties: la formation (70% de l'ensemble de données) et les tests (30% restants) comme indiqué

```
[ ] from sklearn.model_selection import train_test_split
```

```
[ ] X_train,X_test, Y_train, Y_test = train_test_split(X,Y, train_size=0.7,random_state=30)
```

Figure 4.3 Divisez l'ensemble de données en ensemble de formation et de test.

4.2 **Application d'algorithme de données**:

Nous pouvons, afin que nous puissions appliquer notre algorithme de classification, pour chaque algorithme, nous afficherons les différentes fonctions utilisées

```
from sklearn import svm
from sklearn.ensemble import RandomForestClassifier
```

Figure 4.4 – importation des algorithmes

- Nous avons utilisé le formulaire de classification Support Victor Machine (SVM) et Random forest claassifier (RF).

```
[ ]  model = svm.SVC()
```

```
[ ]  model.fit(X_train,Y_train)
     SVC()
```

```
[ ] clf = RandomForestClassifier(n_estimators=100,random_state=0,max_depth=5)
```

```
[ ] clf.fit(X_train,Y_train)
```

La figure 4.5 Déclaration des algorithmes.

5 Les outils de développement :

Dans cette partie nous allons présenter les principaux outils utilisés pour la mise en place de notre projet :

5.1. Python :

LelangagePythonestunlangagedeprogrammationopensourcemulti-plateformes et orienté objet. Grâce à des bibliothèques spécialisées, Python s'utilise pour de nombreuses situations comme le

développement logiciel, l'analyse de données, ou la gestion d'infrastructures. Il n'est donc pas, comme le langage HTML par exemple, uniquement dédié à la programmation web [59].

5.2. Sklearn :

Nous avons utilisé Sklearn qui est une bibliothèque d'apprentissages statique en python, contient toutes les fonctions de l'état de l'art du Machine Learning. On y trouve les algorithmes les plus importants ainsi que diverses fonctions de préprocessing Les fonctionnalités fournies par scikit-learn incluent [60]:

• Régression, (la régression linéaire et logistique).

• Classification, (les voisins les plus proches).

• Clustering, (K-Means ET K-Means++).

• Sélection du modèle.

• Pré-traitement, (la normalisation Min-Max)

Elle set construite a base de :

-Pandas :

Pandas est une excellente bibliothèque pour importer vos tableaux Excel (etautres formats) dans Python dans le but de tirer des statistiques et de charger votre Dataset dans Sklearn [61].

-Seaborn :

Seaborn est une bibliothèque permettant de créer des graphiques statistiques en Python. Elle est basée sur Matplotlib, et s'intègre avec les structures Pandas [61].

5.3. Google Colaboratory :

Colaboratory, souvent raccourci en "Colab", est un produit de Google Research. Colab permet à n'importe qui d'écrire et d'exécuter le code Python de son choix par le biais du navigateur. C'est un environnement particulièrement adapté au machine Learning, à l'analyse de données et à l'éducation. En termes plus techniques, Colab est un service hébergé de notebooks Jupyter qui nenécessi teauc une configuration et permet d'accéder gratuit ementà des ressources informatiques, dont des GPU [62].

6. Conclusions

Dans ce chapitre, nous avons donné un aperçu de notre solution et présenté notre sélection de jeu de données, et pour bien comprendre notre travail, nous avons présenté les étapes que nous avons suivies dans notre projet et les outils utilisés.

Chapitre 5 : Évaluation et Discussion

1 Introduction

Dans ce cinquième chapitre, nous présenterons les résultats de chaque algorithme que nous avons obtenus, ainsi que les déférents méthodes de sélection des attributs. Enfin Nous faisions une comparaison avec les résultats d'une autre étude.

2 Résultat des Algorithmes de machine Learning

À travers le chapitre04, nous avons applique deux algorithmes de Classification (les machines à vecteurs de support, le classificateur de forêt aléatoire).

Les tableaux suivants présentent les paramètres utilisés pour obtenir les meilleurs résultats :

2.1 Application de l'algorithme «le classificateur de forêt aléatoire»

• Résultat des testes :

Algorithme	Paramètre	précision		Recall		F1_Score		Accuracy
		Attack	no attack	Attack	no attack	Attack	no attack	
RF	Randome_state=0 Max_depth=5 N_estimatore=100	0.94	1.00	1.00	0.96	0.97	0.98	0.97

Tableaux 5.1 – L'algorithme « le classificateur de forêt aléatoire »

2.2 Application de l'algorithme «Les machines à vecteurs de support»

• Résultat des testes

Algorithme	Paramètre	précision		Recall		F1_Score		Accuracy
		Attack	no attack	Attack	no attack	Attack	no attack	
SVM	Randome_state=0	0.59	0.73	0.57	0.75	0.58	0.74	0.67

Tableaux 5.2 – L'algorithme « Les machines à vecteurs de support ».

3 Comparaison entre les algorithmes :

Algorithme	Paramètre	précision		Recall		F1_Score		Accuracy
		Attack	no attack	Attack	no attack	Attack	no attack	
RF	Randome_state=0 Max_depth=5 N_estimatore=100	0.94	1.00	1.00	0.96	0.97	0.98	0.97
SVM	Randome_state=0	0.59	0.73	0.57	0.75	0.58	0.74	0.67

Tableaux 5.3 – Comparaison entre les algorithmes

Nous 'avons constaté que l'algorithme "Random Forest Classifier" surpasse l'algorithme "Support Vector Machines" en termes de performances et de précision. Comme indiqué dans le tableau.

algorithme	précision		Recall		F1_Score		Accuracy	précision		Recall		F1_Score		Accuracy
	1	**0**	**1**	**0**	**1**	**0**		**1**	**0**	**1**	**0**	**1**	**0**	
RF	**0.94**	**1.00**	**1.00**	**0.96**	**0.97**	**0.98**	**0.97**	**0.92**	**0.98**	**0.99**	**0.80**	**0.95**	**0.88**	**0.93**
SVM	**0.59**	**0.73**	**0.57**	**0.75**	**0.58**	**0.74**	**0.67**	**0.93**	**0.81**	**0.91**	**0.84**	**0.92**	**0.83**	**0.89**
	nos résultats							**les résultats d'un autre article**						

Tableaux 5.4 – Comparaison entre nos résultats et les résultats d'un autre article.

4 Comparaison entre nos résultats et les résultats d'un autre article :

Nous avons comparé nos résultats final avec les résultats de l'article:«DDoS datasets: Utilisation du Machine Learning pour analyser les performances de détection d'intrusion. D'après les premières observations, on a des bon résultats en comparaison avec le :

_ L'algorithme de classificateur de forêt aléatoire a obtenu de bons résultats par rapport aux autres résultats (Accuracy de 93%, précision de 92 %, Recall de 99 %, respectivement. Notre modèle fonctionne très bien avec une précision F1 de 97 %, un Recall de 100% et une précision de 94 %.

_ Les performances de notre algorithme Les machines à vecteurs de support ne sont pas bonnes par rapport aux résultats de l'autre algorithme Les machines à vecteurs de support , où nous constatons Accuracy , Recall et une précision très médiocres par rapport aux autres résultats.

5 Conclusion

Dans ce chapitre nous avons présenté un résumé de notre travail avec les résultats que nous avons obtenus.

Conclusion Général

Au cours de la dernière décennie, les attaques DDoS sont devenues un important fournisseur de services aux organisations des pays développés et en développement qui n'utilisent qu'Internet et votre ordinateur. Par conséquent, l'objectif général de ces attaques est spécifiquement d'interrompre la disponibilité des services pendant une période prolongée. Cela cause de graves dommages financiers aux organisations. Par conséquent, il est très important de comprendre le problème des attaques DoS et DDoS dans l'environnement.

Dans ce projet, nous avons introduit les différents concepts d'attaques DDoS dans leur environnement, qui est devenu l'un des problèmes de sécurité critiques qui menacent le réseau. Nous avons appris quelques concepts d'apprentissage automatique et également mis en œuvre une solution intelligente pour détecter les attaques DDOS sur différents modèles d'apprentissage automatique en utilisant le langage Python et avons obtenu de bons résultats dans la plupart des algorithmes par rapport à d'autres études.

Références

[1] X. Li, S. Li, J. Hao, Z. Feng, and B. An, “Optimal personalized defense strategy against man-in-the-middle attack,” 31st AAAI Conf. Artif. Intell. AAAI 2017, pp. 593–599, 2017.

[2] X. Liu, H. Liu, Q. Guo, and C. Zhang, Stock prediction model based on wavelet packet transform and improved neural network, vol. 11983 LNCS. 2019.

[3] O. Eigner, P. Kreimel, and P. Tavolato, “Detection of man-in-the-middle attacks on industrial control networks,” Proc. - 2016 Int. Conf. Softw. Secur. Assur. ICSSA 2016, pp. 64–69, 2017, doi: 10.1109/ICSSA.2016.19.

[4] J. Huang, D. W. C. Ho, F. Li, W. Yang, and Y. Tang, “Secure remote state estimation against linear man-in-the-middle attacks using watermarking,” Automatica, vol. 121, p. 109182, 2020, doi: 10.1016/j.automatica.2020.109182.

[5] J. L. Borade, “A Survey on Man in the Middle Attack,” IJSTE - Int. J. Sci. Technol. Eng., vol. 2, no. 09, pp. 277–280, 2016.

[6] Z. Čekerevac, Z. Dvorak, L. Prigoda, and P. Čekerevac, “Internet of Things and the Man-in-the-Middle Attacks – Security and Economic Risks,” MEST J., vol. 5, no. 2, pp. 15–25, 2017, doi: 10.12709/mest.05.05.02.03.

[7] M. Denis, C. Zena, and T. Hayajneh, “Penetration testing: Concepts, attack methods, and defense strategies,” 2016 IEEE Long Isl. Syst. Appl. Technol. Conf. LISAT 2016, 2016, doi: 10.1109/LISAT.2016.7494156.

[8] D. Steinmetzer, Y. Yuan, and M. Hollick, “Beam-stealing: Intercepting the sector sweep to launch man-in-the-middle aacks on wireless ieee 802.11ad networks,” WiSec 2018 - Proc. 11th ACM Conf. Secur. Priv. Wirel. Mob. Networks, pp. 12–22, 2018, doi: 10.1145/3212480.3212499.

[9] B. Pinkas and T. Sander, “Securing passwords against dictionary attacks,” Proc. ACM Conf. Comput. Commun. Secur., pp. 161–170, 2002, doi: 10.1145/586110.586133.

[10] J. Jose, T. T. Tomy, V. Karunakaran, V. Anjali Krishna, A. Varkey, and C. A. Nisha, “Securing passwords from dictionary attack with character-tree,” Proc. 2016 IEEE Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2016, pp. 2301–2307, 2016, doi: 10.1109/WiSPNET.2016.7566553.

[11] A. Sharma, V. Ojha, R. C. Belwal, and G. Agarwal, “Password based authentication: Philosophical survey,” Proc. - 2010 IEEE Int. Conf. Intell. Comput. Intell. Syst. ICIS 2010, vol. 3, pp. 619–622, 2010, doi: 10.1109/ICICISYS.2010.5658405.

[12] A. K. Kyaw, F. Sioquim, and J. Joseph, “Dictionary attack on Wordpress: Security and forensic analysis,” 2015 2nd Int. Conf. Inf. Secur. Cyber Forensics, InfoSec 2015, pp. 158–164, 2016, doi: 10.1109/InfoSec.2015.7435522.

[13] K. Amirtahmasebi, S. R. Jalalinia, and S. Khadem, “A survey of SQL injection defense mechanisms,” Int. Conf. Internet Technol. Secur. Trans. ICITST 2009, 2009, doi: 10.1109/icitst.2009.5402604.

[14] W. G. J. Halfond, J. Viegas, and A. Orso, “A Classification of SQL Injection Attacks and Countermeasures,” Prev. Sql Code Inject. By Comb. Static Runtime Anal., p. 53, 2008.

[15] K. Natarajan and S. Subramani, “Generation of Sql-injection Free Secure Algorithm to Detect and Prevent Sql-Injection Attacks,” Procedia Technol., vol. 4, pp. 790–796, 2012, doi: 10.1016/j.protcy.2012.05.129.

[16] V. K. Gudipati, T. Venna, S. Subburaj, and O. Abuzaghleh, “Advanced automated SQL injection attacks and defensive mechanisms,” 2016 Annu. Connect. Conf. Ind. Electron. Technol. Autom. CT-IETA 2016, 2017, doi: 10.1109/CT-IETA.2016.7868248.

[17] M. Eskandari and S. Hashemi, “A graph mining approach for detecting unknown malwares,” J. Vis. Lang. Comput., vol. 23, no. 3, pp. 154–162, 2012, doi: 10.1016/j.jvlc.2012.02.002.

[18] S. Gadhiya, K. Bhavsar, and P. D. Student, “Techniques for Malware Analysis,” Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 3, no. 4, pp. 2277–128, 2013.

[19] I. A.Saeed, A. Selamat, and A. M. A. Abuagoub, “A Survey on Malware and Malware Detection Systems,” Int. J. Comput. Appl., vol. 67, no. 16, pp. 25–31, 2013, doi: 10.5120/11480-7108.

[20] N. Idika and A. P. Mathur, “A Survey of Malware Detection Techniques,” SERC Tech. Reports, 2007, [Online]. Available: http://www.serc.net/report/tr286.pdf.

[21] M. Khonji, Y. Iraqi, and A. Jones, “Phishing detection: A literature survey,” IEEE Commun. Surv. Tutorials, vol. 15, no. 4, pp. 2091–2121, 2013, doi: 10.1109/SURV.2013.032213.00009.

[22] A. Tewari, A. K. Jain, and B. B. Gupta, “Recent survey of various defense mechanisms against phishing attacks,” J. Inf. Priv. Secur., vol. 12, no. 1, pp. 3–13, 2016, doi: 10.1080/15536548.2016.1139423.

[23] S. Sagar, Shivani, and V. D. Chakravarty, “Phishing attacks and defences,” Int. J. Recent Technol. Eng., vol. 8, no. 1, pp. 894–897, 2019.

[24] https ://images.cigref.fr/Publication/2012-2013-Fondamentaux-Cloud-Computing-Point-de-vue-grandes-entreprises.pdf

[24] https ://softwarelab.org/fr/attaque-ddos/

[25] https ://www.cloudflare.com/fr-fr/learning/ddos/what-is-a-ddos-attack/

[26] https ://doi.org/10.1080/07366981.2018.1453101

[27] https ://www.cloudprotector.com/fr/attaque-ddos

[28] Curt Wilson, Arbor Networks ASERT, « Attack of the Shuriken : Many Hands, Many Weapons ». , fév. 2012.

[29] Brian Krebs, « DDoS Services Advertise Openly, Take PayPal ». , mai 2013.

[30] P. Mockapetris, « Domain names - concepts and facilities ». RFC 1034 (INTERNET STANDARD), nov. 1987. Updated by RFCs 1101, 1183, 1348, 1876, 1982,2065, 2181, 2308, 2535, 4033, 4034, 4035, 4343, 4035, 4592, 5936

[31] D. Mills, J. Martin, J. Burbank et W. Kasch, « Network Time Protocol Version 4 : Protocol and Algorithms Specification ». RFC 5905 (Proposed Standard), juin 2010.

[32] R. Presuhn, « Version 2 of the Protocol Operations for the Simple Network Management Protocol (SNMP) ». RFC 3416 (INTERNET STANDARD), déc. 2002.

[33] Contributing Members of the UPnP Forum, « UPnP™Device Architecture 1.1 »., oct. 2008.

[34] J. Postel, « Character Generator Protocol ». RFC 864 (INTERNET STANDARD),mai 1983.

[35] NTP development team, « NTP Software Downloads ».

[36] Wong Onn Chee, Tom Brennan, «H.....t.....t....p.....p....o....s....t ». , nov. 2010.

[37] S. X. Wu and W. Banzhaf, "The use of computational intelligence in intrusion detection systems: A review," Appl. Soft Comput. J., vol. 10, no. 1, pp. 1–35, 2010, doi: 10.1016/j.asoc.2009.06.019.

[38] A. Patcha and J. M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," Comput. Networks, vol. 51, no. 12, pp. 3448–3470, 2007, doi: 10.1016/j.comnet.2007.02.001.

[39] P. Illy, "Les systèmes de détection d ' intrusion (IDS)," 2018. https://www.researchgate.net/publication/335639245_Les_systemes_de_detection_d'intrusion_IDS (accessed May 27, 2021).

[40] S. M. Othman, F. Mutaher Ba-Alwi, N. T. Alsohybe, and A. T. Zahary, "Survey on Intrusion Detection System Types," Int. J. Cyber-Security Digit. Forensics, vol. 7, no. 4, pp. 444–462, 2018, [Online]. Available: https://www.researchgate.net/publication/329363322.

[41] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in Cloud," J. Netw. Comput. Appl., vol. 36, no. 1, pp. 42–57, 2013, doi: 10.1016/j.jnca.2012.05.003.

[42] S. Sah, "Machine Learning: A Review of Learning Types," 2020. https://www.researchgate.net/publication/342890321_Machine_Learning_

[43] A. Mercier, « L'information face à l'intelligence artificielle : promesses et dangers »https ://larevuedesmedias.ina.fr/linformation-face-lintelligence-artificiellepromesses-et-dangers, 05 février 2018.

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

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

[46] https ://analyticsinsights.io/apprentissage-supervise-vs-non-supervise/

[47] https ://mrmint.fr/naive-bayes-classifier [59] H. B. Barlow. Unsupervised Learning. Neural Computation, vol. 1, no. 3, pages 295-311,1989.

[48] https ://mrmint.fr/naive-bayes-classifier

[49] https ://datascientest.com/knn.

[50] https ://mrmint.fr/introduction-k-nearest-neighbors.

[51] https ://runebook.dev/fr/docs/scikit-learn/modules/svm.

[52] https ://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificie lle/1501879-machine-a-vecteurs-de-support-svm-definition-et-cas-d-usage

[53] https ://scikit-learn.org/stable/modules/svm.html https ://datascientest.com/randomforest-definition

[54] M. Z. Alom et al., "A state-of-the-art survey on deep learning theory and architectures," Electron., vol. 8, no. 3, pp. 1–66, 2019, doi: 10.3390/electronics8030292.

[55] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," J. Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019- 0192-5.

[56] https ://www.analyticsvidhya.com/blog/2020/09/precision-recall-machinelearning/

[57] https ://www.lebigdata.fr/confusion-matrix-definition.

[58] https://data.mendeley.com/datasets/jxpfjc64kr/1

[59] https: //www.futura-sciences.com/tech/definitions/informatique-python-19349/

[60] http ://www.codecademy.com/articles/scikit-learn

[61] Https ://datascientest.com/sea born.

[62]https://research.google.com/colaboratory/faq