République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieure et de la Recherche Scientifique

Université Ahmed Daria – Adrar

Faculté des Sciences et de la technologie

Département des Mathématiques et Informatique

Mémoire de fin d'étude, en vue de l'obtention du diplôme de Master en informatique

Option : Systèmes Intelligents(SI)

Thème

Détection d'intrusions via des réseaux de neurones optimisés par système immunitaire artificiel

Préparé par :

ABBASSI Fatima et KOUIDRI Anwar

Jury:

Encadreur : Mr. SLIMANI Ahmed**.**

Président : Mr. MAMOUNI El Mamoun.

Examinateur : Mr. CHOGUEUR Djilali.

Année Universitaire 2022/2023

الجمهورية الجزائرية الديمقراطية الشعبية

People'sDemocraticRepublic of Algeria

وزارة التعليم العالي و البحث العلمي
جامعة أحمد دراية- أدرار
كلية العلوم والتكنولوجيا
قسم الرياضيات و الاعلام الالي

Ministry of Higher Education and Scientific Research
University Ahmed Draia of Adrar
faculty of science and technology
Departement of Mathematics and Computer Science

شهادة الترخيص بالإيداع

انا الأستاذ(ة): سليماني احمد

المشرف مذكرةالماسترالموسومةبـــ : **Détection d'intrusions via des réseaux de neurones optimisés par système immunitaire artificiel**

من إنجاز الطالب(ة): قويدري أنوار

و الطالب(ة): عباسي فاطمة

كلية : العلوم والتكنولوجيا.

القسم : الرياضيات والاعلام الي

التخصص: اعلام الي

تاريخ تقييم / مناقشة: 2023/06/15.

أشهد ان الطلبة قد قاموا بالتعديلات والتصحيحات المطلوبة من طرف لجنة التقييم / المناقشة، وان المطابقة بين النسخة الورقية والإلكترونية استوفت جميع شروطها.

وبإمكانهم إيداع النسخ الورقية (02) والاليكترونية (PDF).

- امضاء المشرف:

ادرار في : 04 جويلية 2023

مساعد رئيس القسم:

أ. كحيلي محمد
رئيس قسم الرياضيات والإعلام الآلي

قسم الرياضيات والإعلام الآلي

ملاحظة :لاتقبل أي شهادة بدون التوقيع والمصادقة.

REMERCIMENTS

Tout d'abord je remercie le bon dieu tout puissant qui ma guidé dans mes pas, qui ma donné le courage, la volonté, et la patience pour surmonter les épreuves que j'ai rencontré tout au long de ma vie.

Je tiens à exprimer ma sincère gratitude et reconnaissance à mon encadreur :

Mr. SLIMANI Ahmed pour avoir accepté de diriger ce travail, pour son aide, sa disponibilité et ses conseils.

Mes remerciements vont aussi à Monsieur **Mr. MAMOUNI El Mamoun** qui nous a fait l'honneur de présider ce jury.

Je remercie **Mr. CHOGUEUR Djilali** d'avoir accepté d'examiner ce travail.

Enfin je remercie toute personne ayant contribué de prés ou loin à la réalisation de ce travail.

Dédicace

Je dédie ce travail humble

A mes parents, source de tendresse, de noblesse et de générosité.

Mes sœurs et frères, avec mes vœux de bonheur, de santé et de réussite.

Et à toute ma famille « KOUIDRI »

A tous mes amis, tous mes professeurs

Et à tous ceux qui commentent cet humble travail

Anwar

Dédicace

Je dédie ce modeste travail

A tous ceux qui me connaissent, en particulier,

A mon père et ma mère.

A mes frères et mon sœur(Zahra)

Et à toute la famille « ABBASSI »

A tous mes amis et collègues.

A tous ceux qui m'ont soutenu avec amour quand j'étais faible .

Sans oublier tous les professeurs qui ont contribué à ma formation de l'enseignement primaire, moyenne et secondaire jusqu'à l'enseignement supérieur.

Fatima

Introduction général

L'augmentation de la connectivité entre les différents réseaux informatiques et la dépendance croissante à l'égard des systèmes d'information en réseau ont conduit principalement à des besoins en matière de protocoles de sécurité. De plus, cette augmentation s'accompagne naturellement de la croissance du nombre d'utilisateurs aux intentions différentes, qui peuvent parfois être destructrices, rendant les données et les ressources des utilisateurs sensibles et vulnérables au vol et à l'exploitation à des fins malveillantes.

La sécurité des systèmes informatiques vise à protéger l'accès aux données et aux ressources du système ainsi qu'à les manipuler à l'aide d'outils tels que les pare-feux et les logiciels antivirus. Malheureusement, les pare-feu ou les logiciels antivirus sont parfois inefficaces contre ces menaces, qui peuvent se propager rapidement. Pour pallier cette lacune, de nouvelles technologies appelées systèmes de détection d'intrusion sont apparues, dont le but est de détecter les comportements malveillants, qui à leur tour se basent sur l'examen du trafic réseau et la génération de renseignements. Le principal défi des systèmes de détection d'intrusion est d'identifier tout comportement malveillant à l'extérieur ou à l'intérieur du système informatique. Actuellement, le système de détection d'intrusion est un domaine ouvert de recherche et de développement car il repose sur deux approches fondamentales. Tout d'abord, une approche basée sur des scénarios qui repose sur la comparaison du comportement d'utilisation du système avec des signatures d'attaque précédemment connues. Il est interdit de découvrir de nouvelles attaques sans mettre à jour la base de données des signatures. Deuxièmement, l'approche comportementale consiste à construire un nouveau modèle qui identifie les comportements qui s'écartent du comportement normal. Ce modèle est le résultat d'une étape d'apprentissage sur une grande base de données.

Dans cette thèse, nous essayons de réaliser un système de détection d'intrusion comportementale capable de détecter de nouvelles attaques avec un taux de réussite maximal et moins de fausses alarmes. Pour cela nous avons intégré des réseaux de neurones artificiels qui sont un outil d'intelligence artificielle pour l'apprentissage et la classification comme méthode d'optimisation par un système immunitaire artificiel inspiré de l'immunité humaine. Cette combinaison a prouvé ses performances dans de nombreux domaines d'application et dans la résolution de divers problèmes**.**

Organisation de mémoire :

Ce travaille est composé de quatre chapitres .

Après l’introduction générale on trouve le premier chapitre de la sécurité informatique en commençant par définir la sécurité informatique après on définie les terminologies de la sécurité, les menaces et les attaques , ensuite nous parlons sur IDS on définie quelqu’un système de détection d’intrusion on commence par les intrusions , après on focalise sur la détection d’intrusion et on conclue des tests IDS.

Dans le deuxième chapitre on définie les réseaux de neurones artificiels qui représentent un outil d’intelligence artificielle pour l’apprentissage et la classification.

Dans le troisième chapitre on parle des optimisation combinatoire et système immunitaire artificiel.

Dans le dernière chapitre nous montrons les résultats obtenus pour les différentes expérimentales effectues .

Résume

La sécurité des réseaux informatiques est devenue une préoccupation cruciale dans notre société interconnectée et les attaques informatiques sont de plus en plus complexes, ce qui nécessite des techniques de détection avancées pour assurer la protection des systèmes. Dans ce contexte, l'application de réseaux de neurones optimisés par des systèmes immunitaires artificiels se révèle être une approche prometteuse pour la détection des intrusions.

Cette méthode permet de faire face aux attaques sophistiquées en exploitant les capacités d'apprentissage et d'adaptation des réseaux de neurones, renforcées par l'intelligence artificielle. Dans le cadre de ce travail on propose un système de détection d'intrusion basé réseaux de neurones optimisé le système immunitaire artificiel qui s'inspire du système immunitaire naturel en utilisant la base de données KDD99, L'algorithme analyse les connexions réseaux afin de les classer comme normale ou anormales. Les résultats expérimentaux montrent que le système est performant avec un taux de reconnaissance satisfaisant.

Mots clés : Système de détection d'intrusions, systèmes immunitaires naturels, système immunitaire artificial, KDD99.

Abstract

The security of computer networks has become a crucial concern in our interconnected society and computer attacks are becoming increasingly complex, requiring advanced detection techniques to ensure the protection of systems. In this context, the application of neural networks optimized by artificial immune systems is proving to be a promising approach for intrusion detection.

This method makes it possible to cope with sophisticated attacks by exploiting the learning and adaptation capacities of neural networks, reinforced by artificial intelligence. In the framework of this work we propose an intrusion detection system based neural networks optimized the artificial immune system which is inspired by the natural immune system using the database KDD99, The algorithm analyzes the network connections to classify them as normal or abnormal. The experimental results show that the system performs well with a satisfactory recognition rate.

Key words: Intrusion detection system, natural immune systems, artificial immune system.

الملخص:

أصبح أمن شبكات الكمبيوتر مصدر قلق بالغ في مجتمعنا المترابط ، كما أن هجمات الكمبيوتر معقدة بشكل متزايد ، الأمر الذي يتطلب تقنيات كشف متقدمة لضمان حماية الأنظمة. في هذا السياق ، يثبت تطبيق الشبكات العصبية المحسّنة بواسطة أجهزة المناعة الاصطناعية أنه نهج واعد للكشف عن الاختراقات .

هذه الطريقة تجعل من الممكن التعامل مع الهجمات المعقدة من خلال استغلال قدرات التعلم والتكيف للشبكات العصبية ، معززة بالذكاء الاصطناعي. كجزء من هذا العمل ، نقترح نظامًا لاكتشاف التسلل يعتمد على الشبكات العصبية المُحسّنة لجهاز المناعة الاصطناعي المستوحى من نظام المناعة الطبيعي ، باستخدام قاعدة بياناتKDD99 تحلل الخوارزمية اتصالات الشبكة من أجل تصنيفها على أنها طبيعية أو غير طبيعية تظهر النتائج التجريبية أن النظام يعمل بشكل جيد مع معدل التعرف المرضي .

TABLE DES MATIÈRES

LISTE DES TABLEAUX

TABLE DES FIGURES

Introduction

En raison du développement technologique et du développement des moyens et des techniques, le transfert d'informations et de données à travers les réseaux pose un défi opportuniste pour accéder et traiter ces données. Ainsi, la sécurité de l'information joue un rôle majeur à travers un ensemble de technologies et de modèles avancés pour la sécurité de l'accès à l'information. D'autre part, un système de détection d'intrusion est une méthode programmée pour détecter toute tentative de violation d'un problème de sécurité.

Dans ce premier chapitre, nous abordons deux aspects essentiels. La première partie met en lumière les notions fondamentales de la sécurité informatique et des systèmes de détection d'intrusions. Nous commençons par définir les différentes notions de la sécurité informatique, en exposant ses objectifs et la terminologie associée. Ensuite, nous examinons les attaques informatiques, en donnant quelques exemples concrets.

La deuxième partie se concentre sur les systèmes de détection d'intrusions. Nous définissons ces systèmes, expliquons leur principe de fonctionnement, leurs tests et leur efficacité. Enfin, nous discutons des limites et des avantages des systèmes de détection d'intrusions actuels.

À la fin de ce chapitre, nous énumérons quelques limitations et efficacités des systèmes de détection d'intrusions actuellement disponibles.

I- La sécurité informatique

1-Définition

La sécurité informatique vise à protéger l'information contre une large gamme de menaces, de manière à garantir la continuité des transactions, à réduire le plus possible le risque . Cette sécurité est assurée par la mise en œuvre de mesures adaptées, qui regroupent des règles, des processus, des procédures, des structures organisationnelles et des fonctions matérielles et logicielles [1].

2-Principes de sécurité d'un system informatique :

C'est l'ensemble des moyens mis en œuvre pour réduire les menaces qui guettent

un système [2].

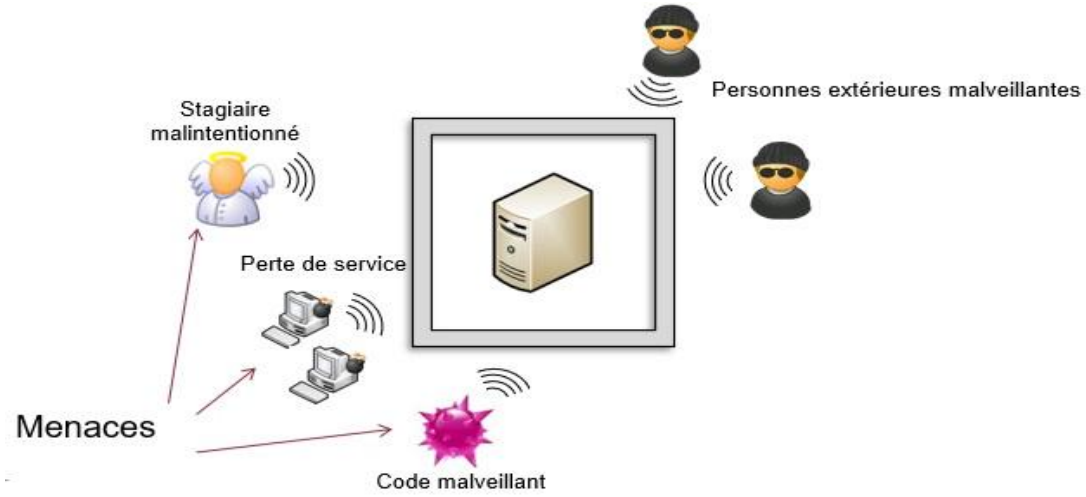


Figure 1.1 -Principes de sécurité d'un system informatique.

Lors de la transmission de messages, différents problèmes peuvent survenir, tant au niveau des échanges entre les parties qu'en présence d'un pirate. Voici quelques exemples de ces problèmes :

Mascarade: il s'agit de l'insertion de messages provenant d'une source frauduleuse dans le réseau.

Modification du contenu : cela englobe toute altération du contenu d'un message, incluant l'insertion, la suppression, la transposition et la modification.

Modification de séquence (ou d'ordre) : il s'agit de toute modification de l'ordre des messages entre les parties, incluant l'insertion, la suppression et la renumérotation.

Modification de la synchronisation : cela implique le retard ou la relecture des messages.

Répudiation de la source : il s'agit du déni de transmission d'un message par la source.

Répudiation de la destination : cela se produit lorsque la destination nie avoir reçu un message[3].

3-Objectifs de la sécurité informatique :

La sécurité est essentielle pour la protection de six caractéristiques critiques des systèmes et de l'information qu'ils traitent et maintiennent à savoir [4]:

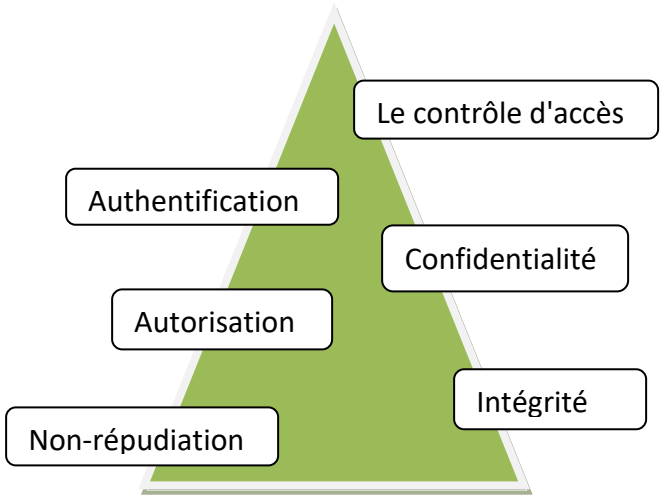


Figure1.2-Objectif de sécurité informatique.

Le contrôle d'accès :limiter et de contrôler l'accès à systèmes et des applications via des maillons de communication et pour obtenir un accès on doit d'abord s'authentifiée, droits d'accès pour chaque cas.

Confidentialité :assure que l'information soit protégée contre toute divulgation accidentelle ou malveillante aux parties non autorisées.

Intégrité : assure que l'information et les systèmes soient accessibles et utilisables par les parties autorisées aux moments où elles en ont besoin .

A coté de ces caractéristiques de bases nous rencontrons également les composants suivantes :

Authentification : assure l'identification d'un individu, d'une entité mais également l'origine de l'information ou encore d'une opération effectuée sur celle-ci.

Autorisation : assure le contrôle de type d'activités ou d'informations qu'une personne ou entité est autorisée à effectuer ou accéder.

Non-répudiation ou irrévocabilité :assure le fait qu'une personne ou entité puisse nier avoir effectué une activité. Dans le domaine du courriel .l'irrévocabilité est utilisée pour assurer que le destinataire ne pourra nier avoir reçu l'information et assurer que l'expéditeur de la source de l'information ne peut nier avoir envoyée l'information .

Au fur et à mesure de votre lecture ,tous ces concepts deviendront plus clairs, du moins tel est le but que fixé [4].

4-Terminologie de la sécurité

4.1 L'actif d'une entreprise :

Un actif est un élément qui doit être protégé et peut inclure une propriété, les personnes et les informations qui ont une valeur pour l'entreprise.

4.2 La classification de l'information :

Une information classifiée est une information sensible dont l'accès est restreint par une loi ou un règlement à un **groupe spécifique de personnes.**

4.2.1 Classification gouvernementale :

-Non classifié
-Sensible mais non classifié
-Confidentiel
-Secret
-Top secret

4.2.2 Classification dans le secteur privé :

-Publique
-Sensible
-Privé
-Confidentiel

4.3 Menace de sécurité :

Est un danger qui existe dans l'environnement d'un système informatique indépendant de celui-ci.

4.4 Vulnérabilité :

C'est une faiblesse ou faille de sécurité au niveau du système informatique qui le rend sensible à une menace.

4.5 Risque :

Est la probabilité qu'une menace particulière puisse exploiter une vulnérabilité donnée du système.

4.6 Contremesures :

C'est-à-dire une manière dont le dispositif ou le processus est mis en œuvre pour contrer une menace potentielle.

4.7 Attaque :

Action malveillante qui tente d'exploiter une faiblesse dans le système et de violer un ou plusieurs propriétés de sécurité.

4.8 Intrusion :

Faute opérationnelle ,externe, intentionnellement nuisible, résultant de l'exploitation d'une vulnérabilité dans le système[5].

5-Menaces et attaques informatique[6][7] :

5.1 Catégorie attaque *:*

Interception : vise la disponibilité des informations (DoS).

Interruption : vise la confidentialité des informations (sniffing, analyse de trafic,...etc).

Modification : vise l'intégrité des informations.

Fabrication **:** vise l'authenticité des Informations.

5.2 Scénarios attaque :

5.2.1 Attaque passive :

Consiste à écouter sans modifier les données ou le fonctionnement du réseau . Elle sont généralement indétectables, mais une prévention est possible et écoutes indiscrètes ou surveillance de transmissions son des attaque de nature passive .Ces attaques passive sont des captures du contenu d'un message et l'analyse de trafic .

5.2.2 Attaque actives :

Ces attaques implique certaines modifier, interrompre et/ou fabrique.il existe une possibilité d'endommagement du systéme,et constitue une menace pour la disponibilité et l'intégrité de l'information d'origine .Dans les attaques actives, les ressources du système peuvent être altérées et l'attention principale de l'attaque est sur la détection.

5.3 Techniques d'attaque :

Il existe plusieurs technique :

5.3.1 Porte dérobée(backdoor) :

Est un point d'entrée dans un système plus ou moins secret, généralement une sécurité pour débloquer un code d'accès perdu ou pour le débogage et aussi un point d'entrée des hackers

5.3.2 Renfilage (sniffing) :

C'est écoute d'une ligne de transmission, cette technique peut être utilisée à l'interne pour le débogage ou de manière abusive par hacker .

5.3.3 Mystification (spoofing) :

C'est une technique d'intrusion qui consiste à envoyer à un serveur des paquets qui semblent provenir d'une adresse IP connue par le pare-feu.

5.3.4 Attaque par rebond(bounce attack) :

Elle menée via un autre ordinateur complice involontaire .Cet ordinateur expédie des messages d'attaque à la victime en masquant l'identité du hacker.

5.3.5 Attaque de l'homme du milieu (man-in-the-middle) :

Se fait passer pour l'un afin d'obtenir le mot de passe de l'autre. Se retourner contre le premier avec un mot de passe valide pour l'attaquer.

5.3.6 Déni de service(denial of service) :

C'est une attaque cherchent à rendre un serveur hors service en le submergeant de trafic inutile.

5.3.7 Trébuchage sans fil(war-driving) :

Dans le cas de réseau sans fil sa consiste à circuler dans la ville avec un portable pour repérer et pénétrer les réseaux locaux non protégés .

5.3.8 Bombe logique :

Est une fonction cachée dans un programme en apparence honnête ,utile ou agréable, qui se déclenchera à retardement .Cette fonction produira alors des actions indésirées,voire nuisibles.

5.3.9 Spamming :

Est une communication électronique il s'agit l'envois en grand quantité effectués à des fins publicitaires ou de déni de service.

5.3.9.1 Spamming canular(hoax) :

Il peut agir courriel de messages publiés sur des forums ou des réseaux sociaux et dont les contenus cherchent à créer l'inquiétude, l'indignation ou au contraire l'approbation.(ex :si tu envoies ce message à 10 personnes, tu aura une bonne nouvelle dans 24 heures).

5.3.9.2 Spamming hameçonnage(phishing) :

C'est une technique fardeleuse consiste a faire croire à la victime qu'elle s'adresse à un tiers de confiance(banque, administration,etc) afin de lui soutirer des renseignements personnels.

5.3.9.3 Spamming

Arnaque :

Technique utilisé par des personnes(escrocs),et promettent la fortune d'un proche, ou celle de l'avocat d'un millionnaire décédé qui permet échange d'un paiement initial.

5.4- Motivation profiles des attaquants (Hacker)

5.4.1 Motivation des hackers :

« hacker » est un mot anglais qui veut dire "bricoleur " ou encore « bidouilleur ».

En informatique ,il est utilisé pour définir les programmeurs débrouillards, avec des connaissances techniques élevées, il sont également capables de détourner un objet ou un logiciel de son fonctionnement originel.

Les hackers sont généralement des personnes cultivées qui connaissent à la fois l'historique de leur statut[7].

5.4.2 Capacités des attaquants :

Transmission de messages sans capacité d'écoute, écoute et transmission de messages, écoute et perturbation des communications , écoute et perturbation et transmissions de messages et relai de messages[7].

5.4.3 Types d'attaquants (les chapeaux noirs(black hats)) :

Ils ne respectent pas la loi ,pénètrent par effraction dans les systèmes dans un intérêt qui n'est pas celui des propriétaires et plus généralement appelés des crackers par exemple créent de virus, de chevaux de Troie ou de logiciels espions[7].

5.4.4 Types d'attaquants (les chapeaux gris(greyhats)) :

Est un peu un hybride des deux précédentes, il s'agit compétent qui agit parfois d'un un chapeau blanc, parfois avec celui d'un chapeau noir ce intention pas mauvaise, mais il comment occasionnellement un délit .Beaucoup de chapeaux blanc s'apparentent en réalité plus à des un chapeau gris[7].

5.4.5 types d'attaquants (les (script kid dies))

Jeunes hackers néophytes, récupèrent les exploits laissés par les white hast .Les exécutent sur des machines, sans aucune connaissance permet de provoquer des pannes volontaires, et jeune adolescent aucune notion de l'éthique d'un hacker[7].

5.5 Les menaces (malware) :

Installés à l'insu des utilisateurs et peuvent générer de nombreux effets indésirables :

Paralysais des performances informatique.

Exploitation des données personnelles.

Suppression des données .

Dysfonctionnement du matériel par ordinateur.

5.5.1 Virus informatiques :

Est la capacité à infecter plusieurs fichiers sur un ordinateur ,il se propagent sur les autres machines lorsque des fichiers infectés sont transférés, la premier virus informatique appelé(brain) en 1986[7].

5.5.2 Vers(Worm) :

Ils ne nécessitent pas d'intervention humaine pour se propager et infecter les ordinateurs ,ils agit un programme capable d'utiliser des réseaux informatique pour infecter les autres machines connectées sans l'aider des utilisateurs et peuvent se répliquer des milliers de fois en vue d'infecter de nouveaux systèmes dans lesquels le processus se reproduira[7].

5.5.3 Adware :

Représente des nuisances les plus couramment rencontrées en ligne, les programmes envoient automatiquement des publicités aux ordinateurs hôtes .Il peut récupérer des informations sur votre site afficher des publicités ciblées sur votre écran ,sont généralement identifiés en tant que (programmes potentiellement indésirables)[7].

5.5.4 Logiciels espions(spaywares) :

Un logiciel espionne ce que vous faites sur votre ordinateur, il recueille des données et les envoyées à des tiers ,généralement des cybercriminels et peut également modifier des paramètres de sécurité spécifique sur votre ordinateur[7].

5.5.5 Ransomware :

Infectent votre ordinateur ,puis chiffrent des données sensibles, puis demandent une rançon pour les récupérer.si vous refusez de payer, les données sont supprimées, certaines variantes de Ransomware verrouillent l'accès à votre ordinateur[7].

5.5.6 Robots :

Sont des programmes conçus pour exécuter automatiquement des opérations spécifiques ,utilisés à de nombreuses fins légales mais ont été redéfinis comme un type de programme

malveillant. Pour fois le robots peuvent exécuter des commandes spécifiques sans que l'utilisateur ne les autorise ou n'en soit informé[7].

5.5.7 Les bugs :

Ne correspondent pas à un type de programme malveillant ,mais à des erreurs commises par un programmeur.il peuvent avoir des conséquences néfastes sur votre ordinateur, blocage ,panne ou réduction des performances.

Permettent aisément aux hackers de passer outre vos défenses et d'infecter votre machine[7].

6-Exemple des attaques informatiques :

Il existe plusieurs exemple[8] :

6.1 Attaque par mot de passe:

L'attaque par mot de passe est une méthode couramment utilisée pour compromettre la sécurité d'un système informatique. Étant donné que les mots de passe sont le mécanisme d'authentification le plus répandu, leur obtention est une approche d'attaque fréquente et efficace. Il existe plusieurs moyens de parvenir à obtenir un mot de passe. Cela peut se faire en fouillant le bureau physique d'une personne pour trouver des indices, en surveillant les connexions au réseau afin d'intercepter des mots de passe non chiffrés, en utilisant des techniques d'ingénierie sociale pour tromper les utilisateurs, en accédant à une base de données de mots de passe compromise, ou tout simplement en devinant le mot de passe. Cette dernière approche peut être effectuée de manière aléatoire ou méthodique en exploitant des informations connues sur la personne ciblée :

- Les attaques par **force brute** , consistent à utiliser une méthode aléatoire en essayant différentes combinaisons de mots de passe, dans l'espoir que l'une d'entre elles fonctionnera. Bien qu'il n'y ait pas de logique spécifique, il est possible d'appliquer une certaine logique en essayant des mots de passe liés au nom de la personne, à son poste, à ses passe-temps ou à des informations similaires..
- Dans une **attaque par dictionnaire**, un dictionnaire contenant des mots de passe couramment utilisés est utilisé pour tenter d'accéder à l'ordinateur ou au réseau d'un utilisateur. Une approche consiste à copier un fichier chiffré contenant les mots de passe, puis à appliquer le même chiffrement à un dictionnaire contenant des mots de passe couramment utilisés, afin de comparer les résultats.

6.2 Attaque par Drive (by Download):

Les attaques par téléchargement furtif, également appelées attaques par Drive-by Download, sont une méthode couramment utilisée pour propager des logiciels malveillants. Les pirates ciblent des sites Web non sécurisés et insèrent un script malveillant dans le code des pages, que ce soit via le protocole HTTP ou PHP. Ce script peut installer des logiciels malveillants directement sur l'ordinateur d'un visiteur du site ou le rediriger vers un site contrôlé par les pirates.

Les téléchargements furtifs peuvent se produire lors de la visite d'un site Web, de l'affichage d'un e-mail ou d'une fenêtre pop-up. Contrairement à de nombreux autres types d'attaques informatiques, un téléchargement furtif ne nécessite pas que l'utilisateur active délibérément l'attaque. Il n'est pas nécessaire de cliquer sur un bouton de téléchargement ou d'ouvrir une pièce jointe malveillante pour être infecté.

Les téléchargements furtifs exploitent souvent des vulnérabilités présentes dans des applications, des systèmes d'exploitation ou des navigateurs Web en raison de mises à jour défaillantes ou d'une absence de mise à jour.

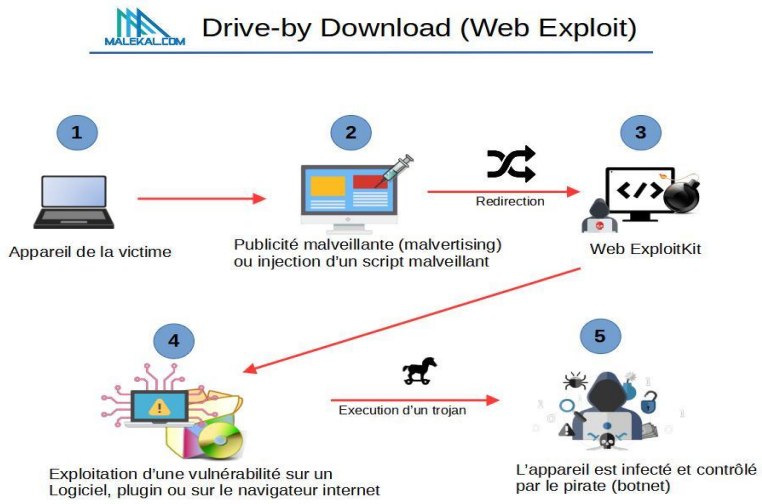


Figure 1.3- Attaque par Drive Download.

6.3 Attaque de l'homme au milieu (MitM)

Une attaque de l'homme du milieu se produit lorsqu'un pirate s'insère dans les communications entre un client et un serveur. Voice quelques examples courante d'attaques de l'homme du milieu:

- Détournement de session : Dans ce type d'attaque, l'attaquant intercepte une session entre un client de confiance et un serveur réseau. L'ordinateur de l'attaquant substitue son adresse IP à celle du client de confiance, tandis que le serveur continue la session en croyant qu'il communique avec le client légitime.

- Attaque par interception : L'attaquant intercepte et écoute les communications entre le client et le serveur, sans modifier les données échangées. Cela permet à l'attaquant d'obtenir des informations sensibles, telles que les identifiants de connexion ou les données confidentielles.
- Attaque par injection : L'attaquant insère ou modifie des données dans les communications entre le client et le serveur. Cela peut inclure l'injection de code malveillant dans les requêtes ou les réponses, ce qui peut entraîner des conséquences néfastes, telles que l'exécution de commandes non autorisées ou la divulgation de données sensibles.

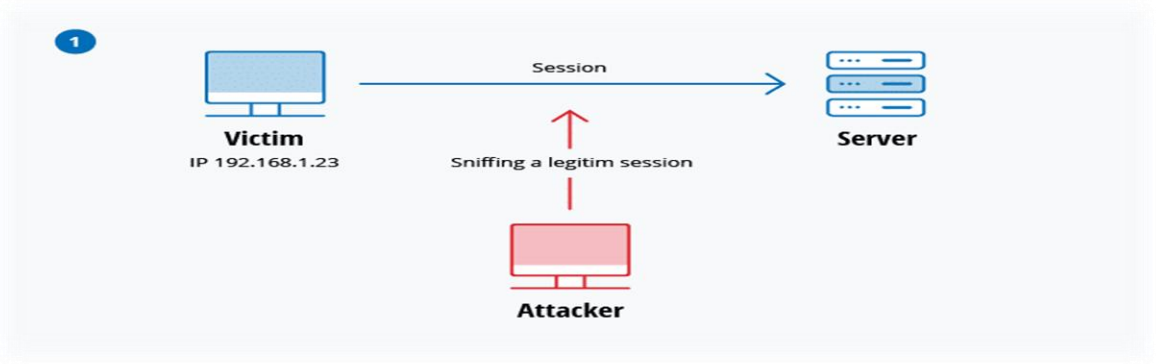


Figure 1.4 - Attaque de l'homme au milieu (MitM).

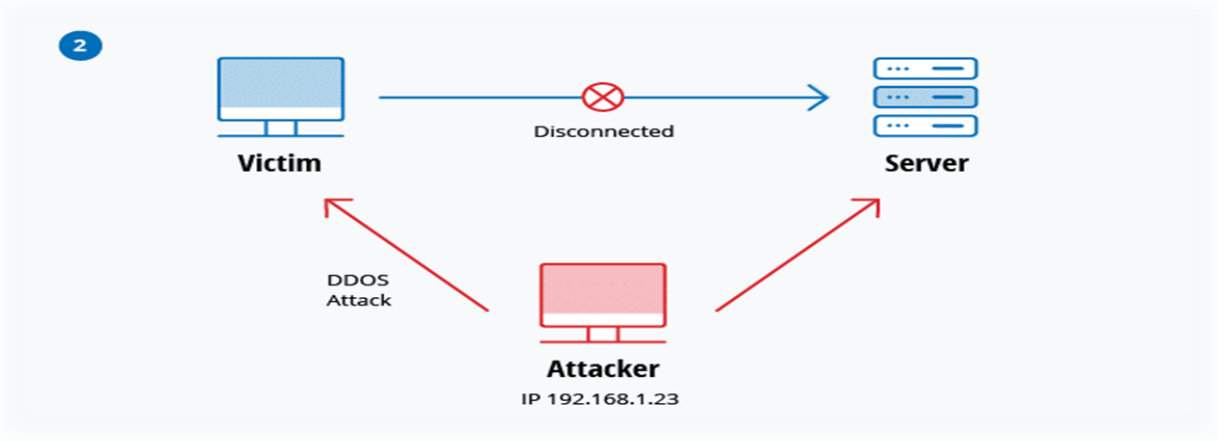


Figure 1.5 -Attaque de l'homme au milieu (MitM).

7-Mécanismes de défense[9] :

Parmi les mécanismes [9] :

7.1 Authentification :consiste à identifier et à vérifier l'association entre l'utilisateur et son identité. L'identité est une information unique associée à l'utilisateur, et connue du système d'authentification et de l'utilisateur. Un service d'authentification repose sur deux composantes :L'identification et l'authentification

7.2 Contrôle d'accès :C'est la gestion des autorisations d'accès à des ressources, une autorisation consiste) gérer et à vérifier les droits d'accès, en fonction des règles de sécurité spécifiées. En dit que un sujet possède un droit d'accès sur un objet si et seulement s'il est autorisé à effectuer la fonction d'accès correspondante sur cet objet.

7.3 Stéganographie : est l'art de cacher un message secret au sein d'un autre message porteur(texte, image, son, vidéo….)de caractère anodin.la sécurité repose sur le fait que la présence même d'un message secret ne sera sans doute pas soupçonnée et détectée

7.3.1 Stéganographie classique : Dissimuler un message secret dans un autre message(ex :collecter les mots de positions impaires),il ne faut pas tuer jules césar car il n'est pas le vrai coupable.

7.3.2 Stéganographie moderne : Si de simuler le message dont :

Les trames→exploiter les champs non-utilisés.

Les exécutables →déclaration des variables strings inutiles

Les images→exploiter les bits de poids faibles du pixel

7.4 Chiffrement : c'est ensemble des techniques permettant de chiffrer des messages c'est-à–dire les rendre incompréhensibles sans une action spécifique. Le message initial est appelé message en clair, après chiffrement message chiffré ou cryptogramme. Le chiffrement et le déchiffrement sont réalisés à partir d'algorithme et utilisé des clés secrètes ou publiques pour réaliser le chiffrement.

7.5 Signature numérique : est le mécanisme de cryptographie utilisé pour vérifier l'authenticité et l'intégrité de données numériques.

7.6 Fonction de hachage : le hachage constitue l'un des éléments principaux d'un système de signature numérique. Tout changement dans les données d'entrée entrainerait une sortie complètement différente.

7.7 Certification :un certification numérique peut être vu comme une carte d'identité numérique, il peuvent servir à l'authentification ,contrôler l'accès à certaines applications ou à certaines données.

7.8 Antivirus : Logiciels conçus pour identifier ,neutraliser et éliminer des logiciels malveillants.il ne protège pas contre un intrus qui emploie un logiciel légitime

II-System de Détection d'intrusion

1-Définition :

Intrusion :

Il s'agit de toute utilisation d'un système informatique dans un but précis autre que celui prévu, c'est-à-dire un ensemble d'actions qui violent la politique de sécurité.

Détection d'intrusion :

La détection d'intrusion est le processus de surveillance des événements qui se produisent dans un système informatique ou un réseau et de leur analyse pour détecter des signes de problèmes de sécurité. Des systèmes de surveillance similaires existent dans d'autres domaines,

Notamment les alarmes antivol et les systèmes de vidéo surveillance que l'on trouve dans les magasins et les banques [10].

2- Classification de system de détection d'intrusion :

Les systèmes de détection d'intrusion sont classés en trois types[11] :

2.1 IDS basé sur l'hôte (HIDS) :

Ce type est placé sur un appareil tel qu'un serveur ou poste de travail, où les données sont analysées localement au machine et collectent ces données à partir de différentes sources. HIDS peut utiliser à la fois un système de détection d'anomalies et d'abus.

2.2 IDS basé sur le réseau (NIDS) :

Les NIDS sont déployés sur des points stratégiques du réseau Infrastructure. Le NIDS peut capturer et analyser des données pour détecter les attaques connues en comparant des modèles ou des signatures de la base de données ou détection d'activités illégales par scan trafic pour activité anormale. Les SNID sont également appelés "renifleurs de paquets", car il capture les paquets qui passent par le biais de supports de communication.

2.3 IDS hybride :

La gestion et l'alerte des dispositifs de détection d'intrusion basés sur le réseau et sur l'hôte, et fournissent la logiquecomplément à NID et HID détection d'intrusion centralisée gestion.

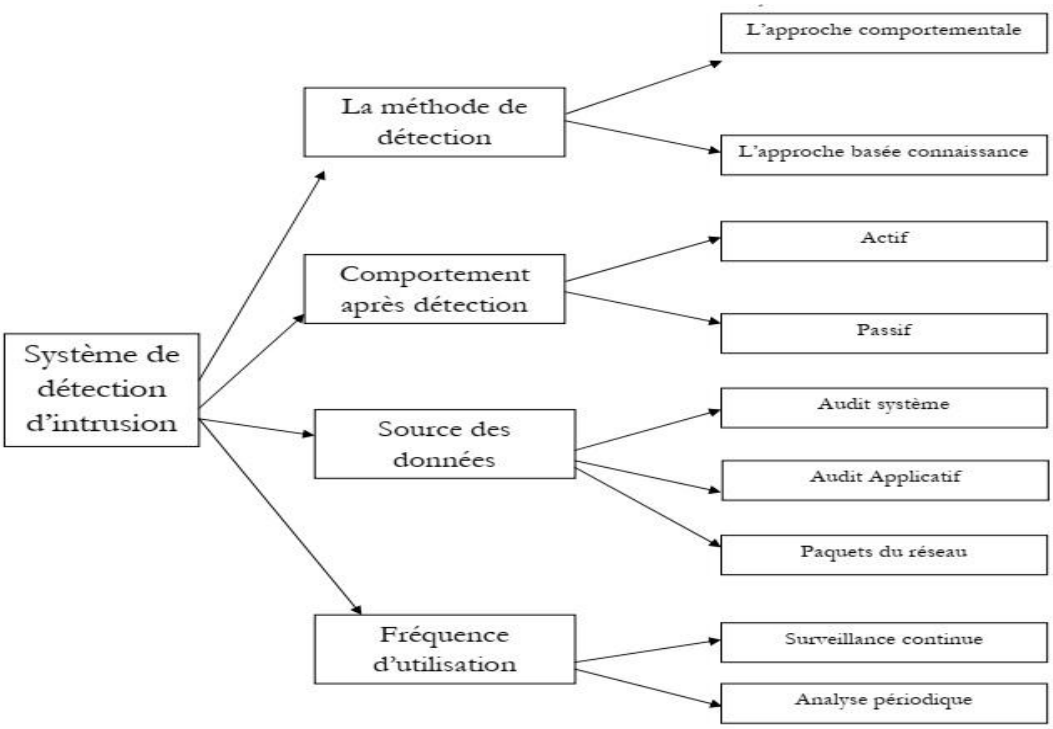


Figure 1.6 -Classification de system de détection d'intrusion.

3-Architecture d'un IDS [12]:

L'architecture d'un IDS est composé de 4 unité .cette architecture est montrée dans la figure suivante :

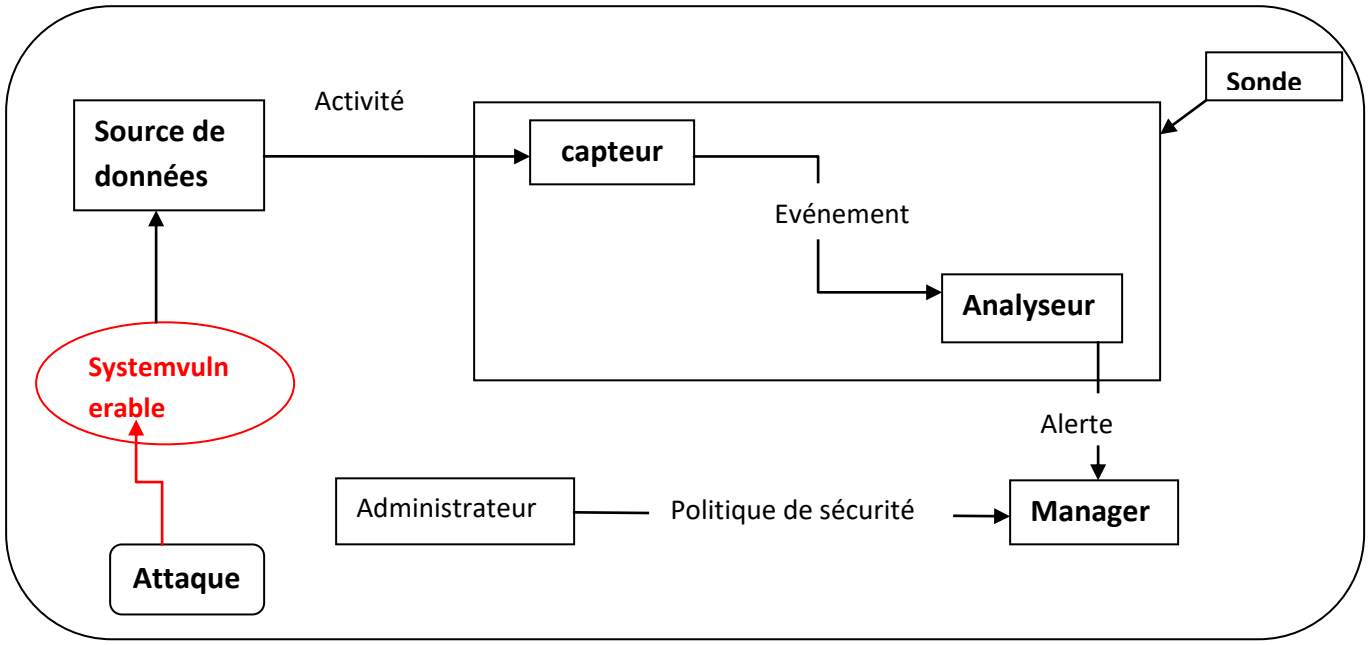


Figure 1.7- Architecture d'un IDS.

Dans l'architecture des systèmes de détection d'intrusions (IDS), différents composants sont impliqués pour assurer la collecte, le traitement et la gestion des informations liées aux activités du système surveillé. Voici une description de ces composants :

1 Source de données : C'est l'interface entre le système surveillé et l'IDS. Elle collecte les informations sur les activités du système et les transmet au reste du système d'IDS.

2 Capteur : Le capteur est chargé de filtrer et de formater les informations brutes provenant de la source de données. Il traite ces informations pour produire un message formaté, appelé événement, qui sera utilisé par la suite dans l'analyse.

3 Analyseur : L'analyseur examine les événements générés par le capteur. Si une activité intrusive est détectée, il émet une alerte, qui est un message standardisé. Le capteur et l'analyseur forment ensemble une sonde, qui constitue un élément clé de l'IDS.

4 Alertes : Lorsqu'une intrusion est détectée par l'IDS, des alertes sont générées. Ces alertes sont généralement stockées dans les journaux du système ou utilisées pour prendre des mesures contre les attaques. Le format standardisé pour formaliser le contenu des alertes est l'IDMEF (Intrusion Détection Message Exchange Format), qui permet une visualisation ultérieure par un expert en sécurité.

5 Manager : Le manager est responsable de la collecte et de la notification des alertes émises par l'analyseur. Il peut également prendre des mesures réactives en cas d'intrusion détectée. Cela peut inclure l'isolement de l'attaque pour réduire les dommages, la suppression de l'attaque, la restauration du système à un état sain et l'identification des problèmes ayant entraîné cette attaque.

4 -Principe de fonctionnement des IDS :

Le fonctionnement d'un IDS et le processus de détection d'intrusion sont illustrés dans la figure suivante :

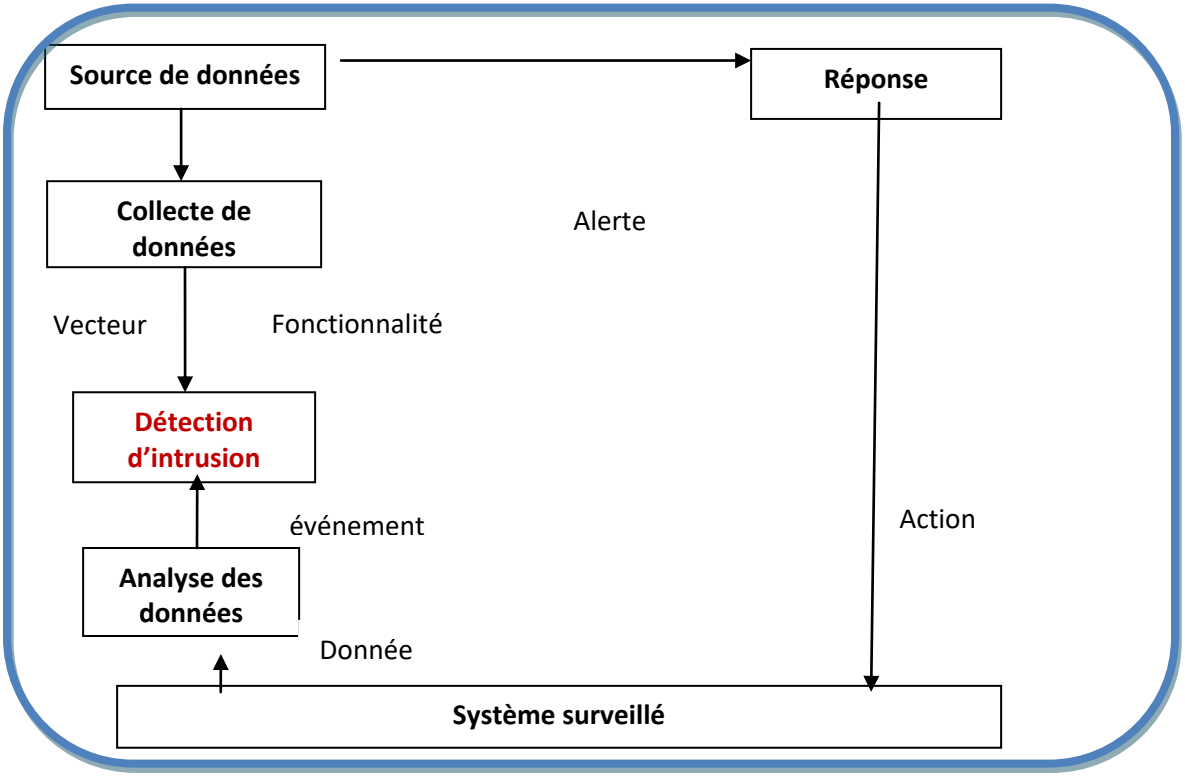


Figure 1.8- Fonctionnement d'un IDS.

5- Emplacement d'un IDS [13]:

Il existe plusieurs endroits stratégiques ou il convient de placer un IDS pour atteindre le niveau de protection attendu selon le plotique se sécurité choisie.Le schéma suivante illustre un réseau local ainsi que les trois positions que peut y prendre un IDS :

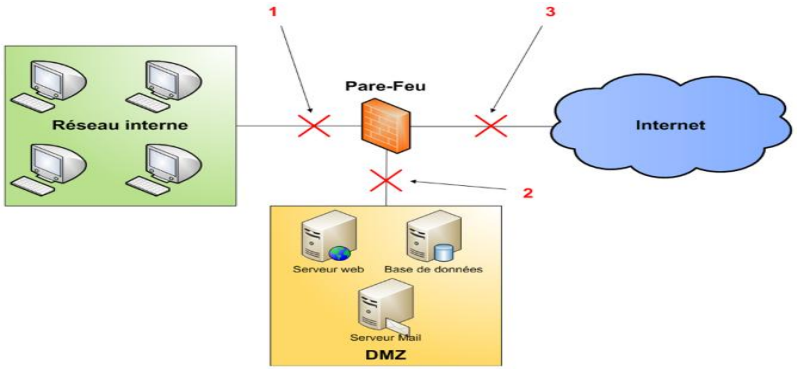


Figure 1.9- Emplacement d'un IDS.

Position 1 : L'IDS dans cette position a pour objectif de rendre compte des attaques internes, provenant du réseau local de l'entreprise. Il peut être judicieux d'en placer un à cet endroit étant donné le fait que 80% des attaques proviennent de l'intérieur.

Position 2 : Si l'IDS est placé sur la DMZ, il détectera les attaques qui n'ont pas été filtrées par le pare-feu et qui relèvent d'un certain niveau de compétence. Les logs seront ici plus clairs à consulter puisque les attaques bénines ne seront pas recensées**.**

Position 3 : Lorsque l'IDS prend cette position, son rôle sera de détecter l'ensemble des attaques frontales, provenant de l'extérieur, vers le pare-feu. Donc, plusieurs alertes seront remontées ce qui rendra les logs difficilement consultables.

6- Efficacité des IDS[14]:

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

Performance : Effectuer une détection en temps réel.

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

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

Complétude : La complétude est la capacité d'un système de détection d'intrusion de détecter toutes les attaques .

7-Les limites des IDS[15]

Les attaquants ont développé des techniques afin de contourner l'IDS , ces techniques peuvent être classées en six catégories :

7.1 L'insertion: cette technique consiste a insérer des données aux flux suspects afin de perturber le fonctionnement de l'IDS.

7.2 La fragmentation: la fragmentation des données peut cacher quelques attaques afin de ne les pas détecter .

7.3 La distribution: c'est la répartition de l'attaque sur plusieurs ressources (attaque DDOS par exemple).

7.4 L'élimination: pénétrer l'IDS et le rendre inutile en le saturant par les flux.

7.5. La substitution: cette technique consiste a échanger le contenu de flux suspect avec son code hexadécimale .

7.6. La confusion: c'est une technique permettant de rendre le contenu incompréhensible.

8 -Tests des IDS :

Avent la mise en place d'un IDS, il est nécessaire de tester ses limites. Pour cela, il existe plusieurs méthodes[16]:

1 Attaque on va utiliser les outils exploités par les attaquants pour détecter une faille dans le système ou dans l'IDS, telles que les techniques d'évasion ou d'insertion.

2 Qualité des informations: on va regarder la qualité des informations fournies par l'IDS lors d'une alarme.

3 Rapidité du système: il est nécessaire que l'IDS soit capable de gérer un grand nombre de données en un temps raisonnable et de détecter l'attaque en un minimum de temps pour réduire les dommages causés.

4 Interaction: le nombre d'interactions entre un IDS et l'administrateur système doit être minime.

Conclusion :

La sécurité doit être un processus transverse à tout système présentant des risques et supportant des menaces. Toutes les méthodologies liées à la systémique s'appliquent dès que nous considérons que les menaces certes exogènes et anachroniques, font partie du première et doivent être surveillées pour traitées .Dans ce chapitre nous avons étudié différentes technologique qui violent la politique de sécurité présenté le système de détection d'intrusion et nous avons également étudié d'une manière détaillée les différentes types d'IDS selon différents critères de classification.

Nous avons examiné en détail l'architecture des systèmes de détection d'intrusions, qui permettent la détection de nouvelles attaques. Pour ce faire, nous nous appuyons sur un système de

détection d'intrusions basé sur des réseaux de neurones, qui sera expliqué en détail dans les chapitres suivants.

Introduction

En raison de l'énorme quantité de données traitées par les ordinateurs ces dernières années, la classification est devenue une méthode utilisée dans de nombreux domaines tels que l'analyse d'images, l'apprentissage automatique, le diagnostic médical...etc.

En sécurité informatique, la classification est utilisée pour catégoriser le trafic réseau ou surveiller les données du système en deux catégories principales (normal/attaque, légal/illégal, etc.). De nos jours, les méthodes de classification sont utilisées dans les systèmes de détection d'intrusion pour prendre des décisions et envoyer des alertes en cas d'attaques, tout en modélisant efficacement le comportement des utilisateurs avec un minimum de fausses alarmes. Pour cela, des méthodes d'intelligence artificielle telles que les réseaux de neurones peuvent être utilisées.

I-Classification :

1-Définition :

Classification correspond aux situations dans lesquelles la machine doit prédire la valeur d'une variable qualitative (variable discrète).Autrement dit la machine doit classer ce qu'on lui donné dans les classes [17].

Les classificateurs, par définition, sont la fonction d'assigner une ligne ou une étiquette de classe à un ensemble fini deBalises pour l'élément en cours de catégorisation. Par exemple, nous avons des informations sur un article Informatif, l'auteur peut préciser le sujet qu'il traite (économique, politique, etc.).Nous avons une image d'un alphabet, le classificateur peut décider quelle lettre de l'alphabet. Classeurs ,il peut être construit automatiquement ou il peut être construit par des experts humains. en fonction de sa structure, Les classeurs générés automatiquement peuvent être interprétés par des humains (par exemple, règles de classification, arbres de décision)[18].

2-L'objectif de classification :

L'objectif principal de la classification supervisée est de définir des régles permettant de Classer des objets dans des catégories en utilisant des variables qualitatives ou quantitatives qui caractérisent ces objets . Les méthodes de classification peuvent également être étendues pour traiter des variables quantitatives (régression)[19].

3-L'architecteur typique d'une application basée sur la classification :

La classification est une phase important de data ming pour extraction les données. L'architecteur présenter dans la figure suivant[20] :

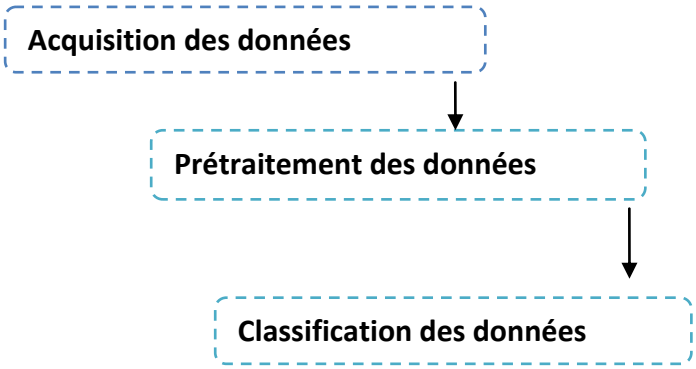


Figure 2.1- Processus de data ming.

Acquisition des données :

Dans ce cas, il est nécessaire de mettre en place un ensemble d'instruments tels que des capteurs et du matériel d'acquisition afin de reproduire le phénomène observé de manière aussi fidèle que possible. Cela implique la transformation de l'information à traiter en signaux numériques manipulables par ordinateur, ce qu'on appelle la numérisation de l'information. Dans notre cas, cela implique la mise en place de sondes (IDS) pour surveiller le trafic réseau.

Prétraitement des données :

Dans cette phase, on effectue le filtrage des informations en ne conservant que ce qui est pertinent dans le contexte de l'étude. Les données collectées peuvent contenir différents types d'anomalies, telles que des erreurs de frappe ou des erreurs système, ce qui nécessite leur traitement. Il peut également y avoir des incohérences dans les données, ce qui nécessite leur exclusion ou leur normalisation. Parfois, il est nécessaire d'effectuer des transformations sur les données afin d'unifier leur poids ou leur format.

Classification des données :

Cette étape correspond à la phase de décision où l'on doit choisir la technique appropriée pour extraire les connaissances des données. Parmi les techniques couramment utilisées, on trouve les réseaux de neurones, les arbres de décision, les réseaux bayésiens, etc. Dans notre cas, la classification TCP/IP se base sur l'utilisation des réseaux de neurones.

4-Catégorie de classification :

Il existe de nombreuses méthodes pour résoudre les problèmes de classification. Cependant, il est possible de les regrouper hiérarchiquement, car certaines approches partagent des caractéristiques communes, que ce soit dans le type d'apprentissage utilisé (supervisé ou non) ou dans la nature de la sortie produite (groupes disjoints ou classification flou)[21].

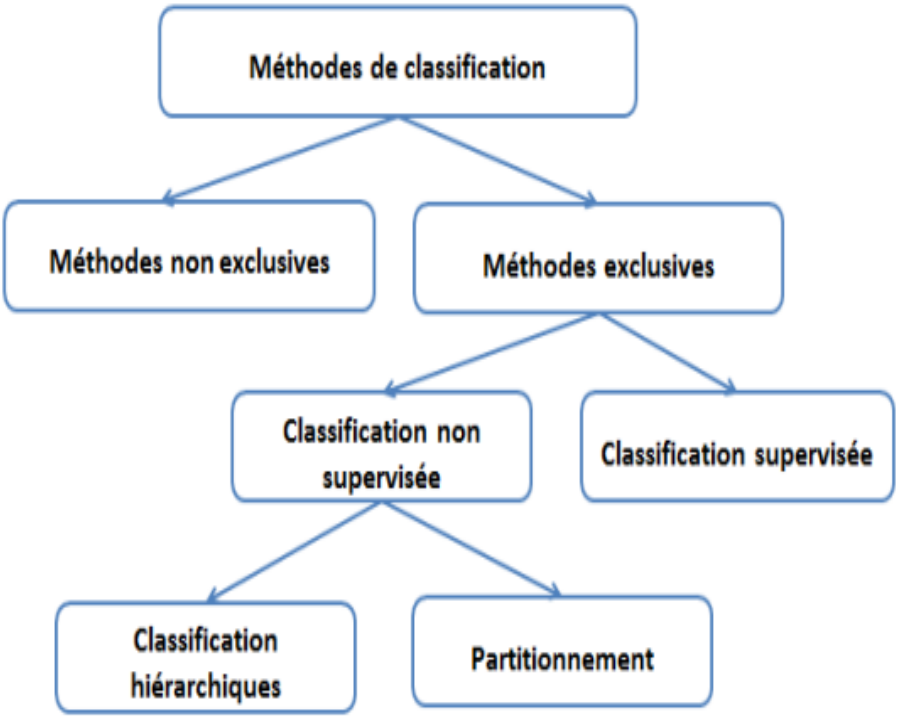


Figure 2.2– Les méthodes de classification.

4.1 Classification exclusive :

Dans la classification exclusive, un objet ne peut appartenir qu'à une seule classe dans la partition finale.

4.1.1 La classification non-supervisée :

La classification cherche à trouver la partition la plus naturelle des données. Les méthodes hiérarchiques, dont certaines sont présentées ci-après, génèrent une séquence de partitions imbriquées. En revanche, les méthodes de partitionnement ne produisent qu'une seule partition des données[22].

4.1.1.1 Classification ascendante hiérarchique (CAH)[23] :

La classification ascendante hiérarchique a pour objectif de construire une série de partitions emboîtées des données en n classes, n-1 classes, ..., 1 classe.

Principe :

1-Au début, la dissemblance entre les N objets est calculée.

2-Ensuite, on regroupe les deux objets dont le regroupement minimise un critère d'agrégation donné, créant ainsi une classe comprenant ces deux objets.

3-On calcule ensuite la dissemblance entre cette classe nouvellement formée et les N-2 autres objets en utilisant le critère d'agrégation. Puis on regroupe les deux objets ou classes d'objets dont le regroupement minimise le critère d'agrégation.

4.1.1.2 Partitionnement :

K-moyennes: L'algorithme des K-moyennes (K-means) est populaire et largement connu dans les méthodes de regroupement. Cet algorithme effectue une partition stricte ("dure"), ce qui signifie que chaque objet est assigné à une seule classe. Il s'agit d'une procédure simple et itérative dont l'idée générale est de regrouper un ensemble X = {x1,...,xn) d'éléments en un nombre fixé à l'avance de K groupes (clusters).

4.1.2 La classification supervisée :L'apprentissage supervisé s'appuie sur des informations fournies par un expert pour apprendre les caractéristiques propres à chaque groupe. L'objectif final est de pouvoir reconnaître automatiquement la catégorie à laquelle appartient un nouvel individu en fonction de ses caractéristiques.

4.1.2.1 K plus proches voisins : Les techniques statistiques "modernes" sont considérées comme des procédures de classification non paramétriques, ce qui signifie qu'elles peuvent être utilisées sans faire d'hypothèses sur l'existence d'une loi de probabilité sous-jacente. La méthode des k plus proches voisins en est un exemple. L'idée générale est la suivante :

Pour un nouvel objet x, on détermine la classe de chacun des k individus les plus proches de x parmi tous les objets de l'ensemble d'apprentissage. Ensuite, on classe le nouvel objet dans la classe la plus fréquente parmi celles de ses voisins.[23][24].

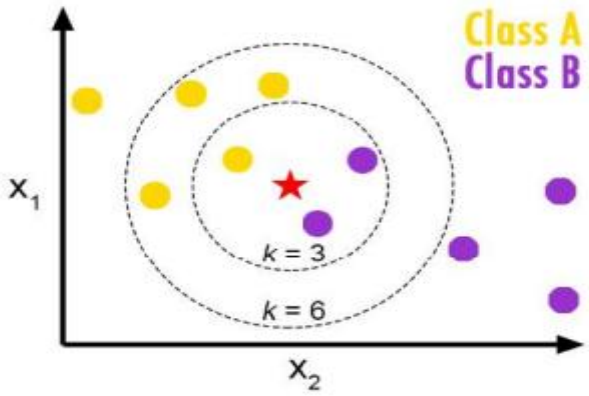


Figure 2.3 – Pour k = 3 la classe majoritaire du point central est la classe B, mais si on change la valeur du voisinage k = 6 la classe majoritaire devient la classe A.

4.1.2.2 Les arbres de décision [5][6][25] : Les arbres de décision sont utilisés pour classer une population d'individus en fonction des valeurs de leurs attributs. Ils représentent graphiquement la procédure de classification, où chaque feuille représente une classe, chaque nœud spécifie un test à effectuer sur un attribut spécifique, et chaque branche correspond à une valeur possible de cet attribut.

Structure d'un arbre de décision[7] : Chaque nœud interne d'un arbre de décision est responsable de la répartition homogène des éléments à classifier parmi ses fils, en se basant sur une variable discriminante spécifique de ces éléments. Les branches qui relient un nœud à ses fils représentent les différentes valeurs de la variable discriminante du nœud. Enfin, les feuilles de l'arbre représentent les résultats de la prédiction pour les données à classifier.

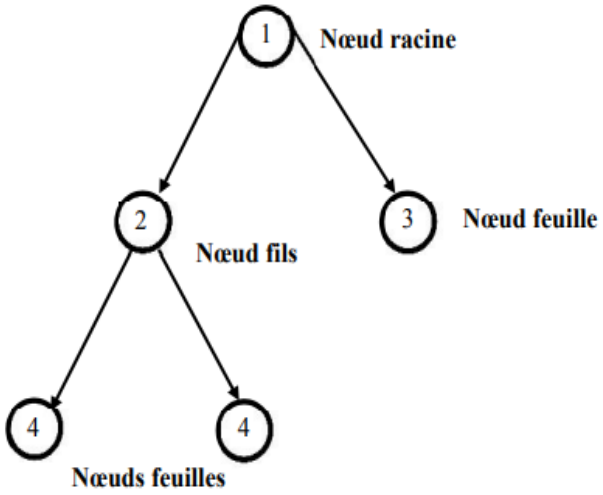


Figure 2.4 -Structure d'un arbre de décision.

4.2 Classification non exclusive:

Dans la classification floue, chaque objet est associé à une densité de probabilité qui indique, pour chaque classe, la probabilité que l'objet considéré appartienne à cette classe. Cela permet de prendre en compte le degré d'appartenance d'un objet à chaque classe plutôt qu'une assignation binaire[26].

II-Réseaux de neurones:

1-Définition :

L'ensemble des neurones forme un graphe pondéré sur lequel circule un signal généré par des stimuli extérieurs (les entrées).

La réseau de neurones est composé de nœuds connectés les uns aux autres , ou chaque nœud correspond 1 à une variable.

Un réseau neuronal est une association d'objets élémentaires, les neurones formels, dans un graphe plus ou moins complexe. Les différents réseaux se distinguent par l'organisation du graphe (couches, connexions complètes, etc.), c'est-à-dire leur architecture, leur niveau de

complexité (nombre de neurones, présence de boucles de rétroaction), le type de neurones (fonctions de transition ou d'activation) et l'objectif recherché : apprentissage supervisé ou non supervisé, optimisation, systèmes dynamiques, etc[27].

2-Les principales composants de réseaux de neurone [28] :

- **Neurones :ensemble des fonctions :**

Ils prennent une donnée d'entrée et produisent une donnée de sortie. Un certain nombre de neurones sont groupés en couches (ou layer). Tous les neurones du même groupe remplissent un type de fonction similaire.

Les neurones d'entrée reçoivent des données d'entrée, les traitent et les transmettent aux neurones dans la couche suivante. Les neurones cachés prennent les données de sortie des précédents neurones en entrée, calculent de nouvelles données de sortie et les transmettent à des couches successives.

- **Couches :groupement des neurones**

Les couches (ou layer) contiennent des neurones et aident à faire circuler l'information. Il existe au moins deux couches dans un réseau de neurones: la couche d'entrée (input layer) et la couche de sortie (output layer).

- **Poids et biais :valeur numérique**

Les poids et biais sont des variables du modèle qui sont mises à jour pour améliorer la précision du réseau. Un poids est appliqué à l'entrée de chacun des neurones pour calculer une donnée de sortie.

Les réseaux de neurones mettent à jour ces poids de manière continue. Il existe donc une boucle de rétroaction mise en œuvre dans la plupart des réseaux de neurones.

- **Fonction d'activation :algorithme mathématique appliqués aux valeur de sortie**

Les fonctions d'activation lissent ou normalisent la donnée de sortie avant qu'elle ne soit transmise aux neurones suivants. Ces fonctions aident les réseaux de neurones à apprendre et à s'améliorer.

3-Neurone formel[29] :

La première modélisation d'un neurone a été présentée par Mac Culloch et Pitts en 1943. Ils ont proposé le modèle suivant : « Le neurone formel fait une somme pondérée des potentiels d'activation x1,x2,...,en qui lui parviennent, puis s'active suivant la valeur de cette sommation pondérée. Si cette somme dépasse un certain seuil, le neurone est activé et transmet une réponse, si le neurone n'est pas activé il ne transmet rien » Ce modèle est schématisé par la figure suivante:

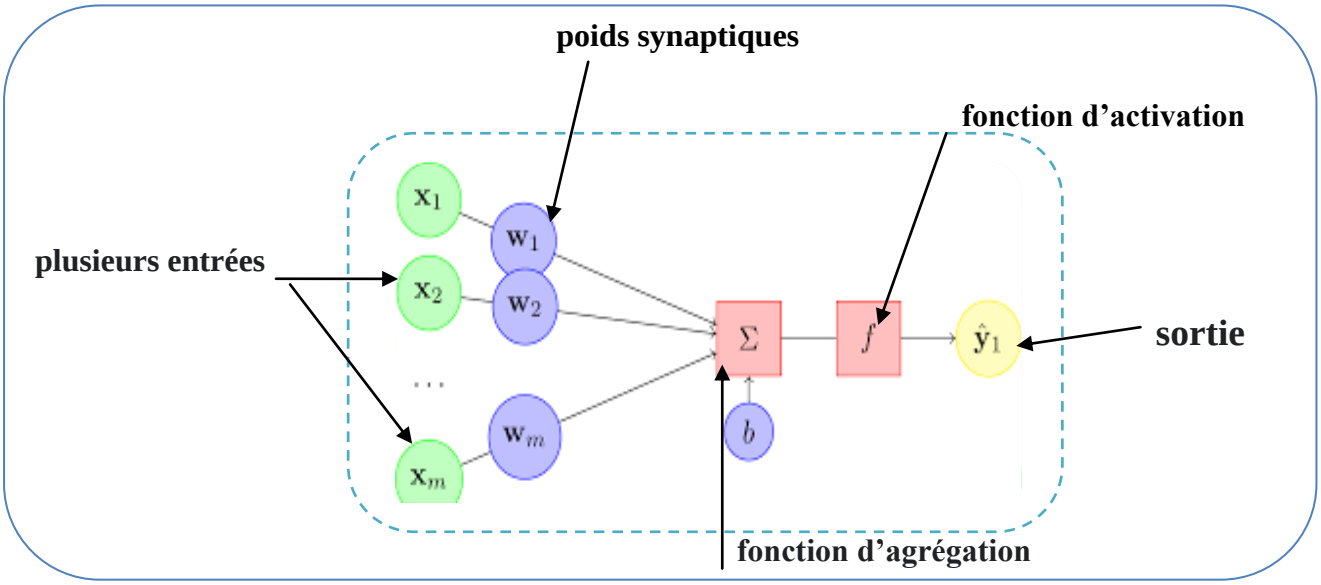


Figure 2.5 -Structure d'un neurone formel.

4-Neurone biologie :

Le neurone biologique est une cellule vivante spécialisée dans le traitement des signaux électriques. Les neurones sont interconnectés par des liaisons appelées axones. Ces axones jouent un rôle crucial dans le comportement logique global du système. Les signaux électriques sont transmis le long des axones, de la sortie d'un neurone vers l'entrée d'un autre neurone, à travers des synapses. Les neurones effectuent une sommation des signaux d'entrée et en fonction du résultat, génèrent un courant en sortie[30].

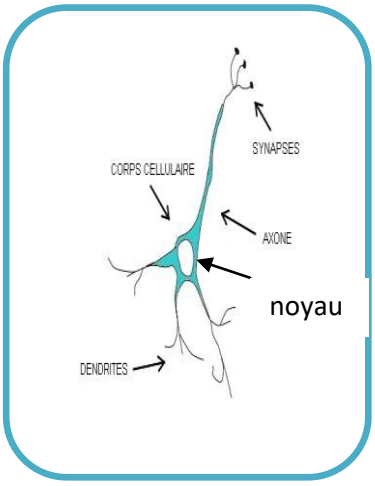


Figure 2.6 -Model d'un neurone biologie.

5-Modélisation des neurone formel :

La modélisation se situe au niveau d'un système de réseau de neurones artificiels selon le principe biologique, qui à son tour correspond à chaque élément constituant le neurone biologique. Dans ce tableau, nous pouvons résumer la transition entreles neurones biologiques et les neurones formels [30]

Neurone biologique	Neurone formel
Synapses	Poids des connexions
Axones	Signal de sortie
Dendrites	Signal d'entrée
Noyau	Fonction d'activation

Tableaux 2.1- Analogie entre les neurones biologiques et les neurones formels.

6-Architecteur des réseaux de neurones [31]:

6.1 Réseaux bouclés : les neurones ne peuvent pas être ordonnés de sorte qu'il n'y ai pas de connexion vers l'arrière.

6.2 Réseaux non bouclés (réseaux à couches) : les neurones peuvent être ordonnés de sorte qu'il n'y ai pas de connexion vers l'arrière[31] .

Exemple :réseaux à couche intermédiaire

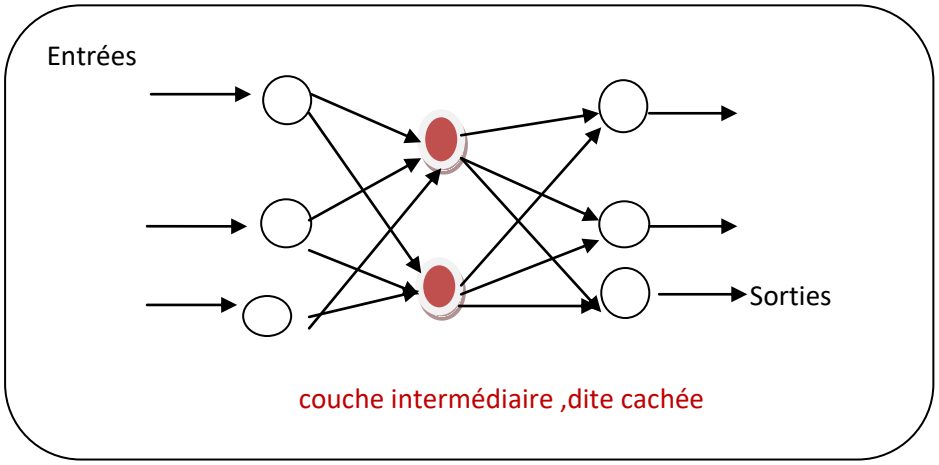


Figure 2.8-Réseaux de neurones non bouclés.

Si Y est le vecteur des sorties et X le vecteur des entrées :

Y=f(WX)

f :fonction d'activation du réseaux

W :vecteur des « poids » des liaisons synaptiques

Apprentissage = détermination des poids permettant d'obtenir une sortie proche d'une sortie Y0 voulue à partir d'une entrée X.

7-Modèle des réseaux de neurones :

7.1 Modèle de Perceptron multi couche :

Le perceptron est un type couramment utilisé de réseau de neurones, adapté pour des problèmes d'approximation, de classification et de prédiction. Il est composé d'au moins trois couches de nœuds : une couche d'entrée, une couche cachée et une couche de sortie. À l'exception des nœuds d'entrée, chaque nœud est un neurone qui utilise une fonction d'activation non linéaire. Ce modèle est également utilisé dans les systèmes de détection d'intrusion. La figure suivante illustre un exemple de formulaire de perceptron[26] :

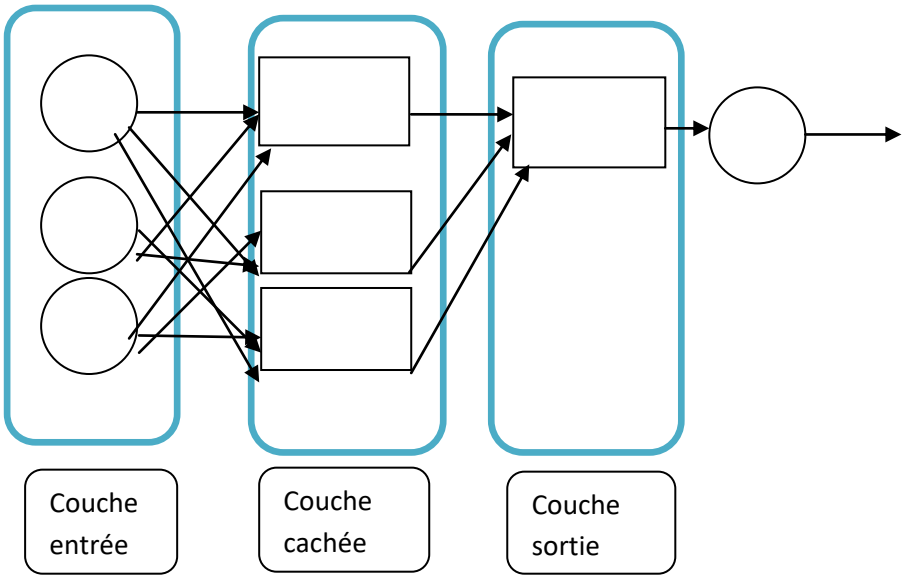


Figure 2.9 -Modèle de perceptron multi couche.

7.2 Modèle de Kohonen :

Ce modèle a été présenté par T. Kohonen en 1982 en se basant sur des constatations biologiques,les cartes auto-organisatrices de Kohonen sont constituées un réseaux .Il a pour objectif de présenter des données complexes et appartenant généralement à un espace discret de grandes dimensions.Les neurones de la couche de sortie sont placés dans un espace d'une ou de deux dimensions. Chaque neurone de la couche de sortie possède des connexions latérales récurrentes dans sa couche[26].

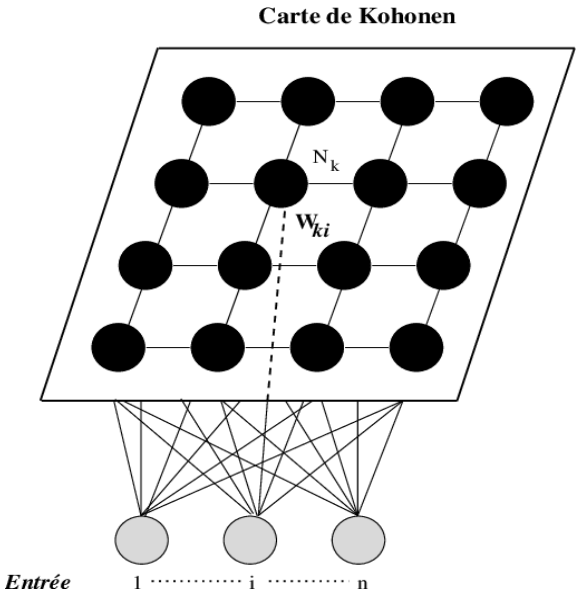


Figure 2.10 - Modèle de kohonen.

7.3 Modèle de Hopfield :

Le modèle de Hopfield fut présenté en 1982.il apparaît Hopfield a montré que un grand nombre de neurones hautement stylisés ont des propriétés collectives. Il a trouvé qu'un ensemble de fonctionnement des neurones non linéaires peuvent stocker des informations avec stabilité et efficacité, rappelez-le avec quelques corrections d'erreurs capacité, et faire preuve d'un sens de l'ordre du temps. Aussi, son modèle est assez robuste et devrait fonctionner même lorsque plus des détails neurologiques sont ajoutés[26].

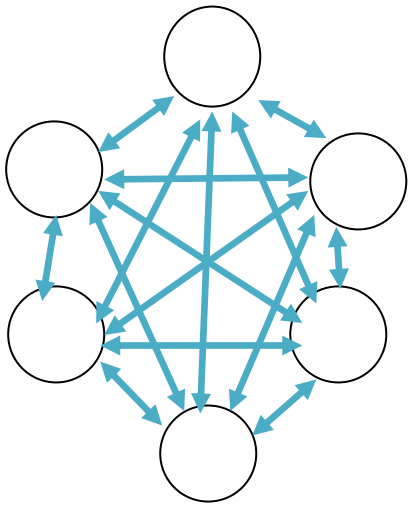


Figure 2.11-Modèle de Hopfield.

8-Apprentissage des réseaux de neurones :

- Mise à jour des poids de connexions ,en général à partir d'un ensemble de données d'entrainement.
- Modification itérative des poids
- Paradigme d'apprentissage :modélisation de l'environnement dans lesquels le réseaux opérera
- La formation des réseaux de neurones est le processus d'apprentissage des réseaux de neurones pour effectuer une tâche. Les réseaux de neurones apprennent en traitant plusieurs grands ensembles de données initialement nommées ou non étiquetées. Ensuite, il peut traiter les entrées inconnues avec plus de précision en utilisant ces exemples ,il existe 3 types d'apprentissage est [32]:

Apprentissage supervisée :

Les méthodes d'apprentissage supervisé nécessitent que la valeur de la variable de sortie pour chaque échantillon d'apprentissage soit connue. Par conséquent, chaque échantillon d'apprentissage se présente sous la forme d'une paire de valeurs d'entrée et de sortie. L'algorithme entraîne ensuite un modèle qui prédit la valeur de la variables de sortie à partir des variables d'entrée à l'aide des caractéristiques définies dans le processus[32].

Apprentissage non supervisée :

Les techniques d'apprentissage non supervisé ne nécessitent que des fonctions d'entrée valeurs dans les données d'entraînement et l'algorithme d'entraînement découvre une structure cachée dans les données d'entraînement basées sur celles-ci[32].

Apprentissage par renforcement :

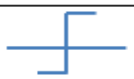
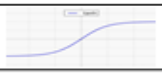
L'apprentissage demandant un grand nombre d'expérimentation, il est très utile voir indispensable de disposer d'un environnement simulé[32].

9-Fonctionnement des réseaux de neurones :

Pour créer une machine capable d'apprendre, les premiers informaticiens ont fait beaucoup de recherches de preuves en étudiant d'autres choses qui sont bonnes pour apprendre et il s'avère qu'il n'y a rien d'aussi efficace et de meilleur pour apprendre que le cerveau humain. Notre cerveau est composé de cellules très particulières appelées neurones. Les neurones ont deux extrémités : les signaux d'entrée entrent à une extrémité et sont combinés ensemble à l'intérieur du neurone et quittent l'autre extrémité en une seule sortie. Tous les milliards de neurones sont connectés les uns aux autres, dans ce que nous appelons un réseau de neurones biologiques.

C'est ainsi que notre cerveau traite les informations et reconnaît les modèles. Les premiers informaticiens ont donc décidé d'imiter les neurones humains en créant leurs propres réseaux de neurones artificiels. Rien d'inhabituel, juste plusieurs signaux entrant en tant qu'entrées et traversant les neurones, ils y sont combinés et traités à l'aide de calculs simples, et un nouveau signal émerge. C'est un bon début, mais un neurone ne fait pas grand-chose. Le plein potentiel

de ce système est révélé lorsque les neurones sont connectés pour former un réseau neuronal artificiel,et c'est ce qui permet aux ordinateurs de reconnaître la sortie. Le tableau suivante montre différente fonction de réseaux de neurone [33]:

Nom de la fonction	*Relation d'entrée/sortie*	*Icône*
Seuil	a=0 si n<0 a=1 si n≥0	
Seuil symétrique	a=-1 si n<0 a=1 si n≥0	
Linéaire	a=n	
Linéaire saturée	a=0 si n<0 a=n si 0≤n≤1 a=1 si n>1	
Linéaire saturée symétrique	a=-1 si n<-1 a=n si -1≤n≤1 a=1 si n>1	
Linéaire positive	a=0 si n<0 a=n si n≥0	
Sigmoïde	$a=\frac{1}{1+e^{-n}}$	
Tangente hyperbolique	$a=\frac{e^{n}-e^{-n}}{e^{n}+e^{-n}}$	
compétitive	a=0 si n maximum a=1 autrement	ℂ

Tableaux 2.2- Différentes fonctions d'activations utilisées dans les RNA[19].

10-Perceptron multi couche MLP :

Le perceptron multicouche (MLP) est une extension du réseau de neurones à propagation avant. Il est composé de trois types de couches : la couche d'entrée, la couche de sortie et les couches cachées, comme illustré dans la Figure 3. La couche d'entrée reçoit le signal d'entrée à traiter. La tâche requise, telle que la prédiction ou la classification, est effectuée par la couche de sortie. Les couches cachées, situées entre les couches d'entrée et de sortie, constituent le moteur de calcul principal du MLP. Comme dans un réseau à propagation avant, les données circulent de la couche d'entrée à la couche de sortie. Les neurones du MLP sont entraînés à l'aide de l'algorithme d'apprentissage par rétropropagation. Les MLP sont conçus pour approximer n'importe quelle fonction continue et peuvent résoudre des problèmes qui ne sont pas linéairement séparables. Les principales applications du MLP sont la classification, la reconnaissance, la prédiction et l'approximation de modèles[19].

10.1 Mise en œuvre du réseau de neurones MLP :

La mise en œuvre des réseaux de neurones comprend à la fois une phase de conception visant à choisir la meilleure architecture possible, et une phase de calcul numérique pour réaliser l'apprentissage du réseau. Pour améliorer le fonctionnement du MLP et réduire le temps de calcul, il est important de rechercher une architecture optimale en termes de nombre de couches, de nombre de neurones par couche et de nombre de sorties possibles.

À partir d'une architecture de réseau de neurones donnée et des exemples disponibles dans la base d'apprentissage, on détermine les poids optimaux en utilisant l'algorithme de rétropropagation des erreurs. L'objectif est d'ajuster les poids de manière à ce que la sortie du modèle se rapproche autant que possible du comportement désiré. Cela permet d'optimiser la performance du réseau de neurones et d'obtenir des résultats précis lors de la classification, de la prédiction ou de l'approximation de modèles[19].

10.2 L'apprentissage des réseaux MLP[19] :

L'apprentissage d'un réseau de neurones multicouches est généralement réalisé à l'aide de l'algorithme de rétropropagation, également connu sous le nom de "back propagation". C'est l'algorithme d'apprentissage supervisé le plus couramment utilisé. L'idée principale de cet algorithme est de calculer le gradient de l'erreur pour chaque neurone du réseau, en remontant de la dernière couche vers la première.
Le processus d'apprentissage par rétro propagation peut être décrit en trois étapes principales :

- Acheminement de l'information à travers le réseau.
- Rétro propagation des sensibilités et calcul du gradient .
- Ajustement des paramètres par la règle du gradient approximé.

Propagation avant : Dans l'étape de rétropropagation, on calcule d'abord la valeur d'entrée pour chaque neurone j de la couche cachée ou de la couche de sortie. Cette valeur est obtenue en faisant la somme pondérée des valeurs de sortie des neurones de la couche précédente. La formule de calcul est la suivante :

$$VEj = \Sigma i \in pred(j)\ Wij * VAi$$

Ensuite, on calcule la valeur d'activation de chaque neurone j en utilisant une fonction de transfert f. Une fonction d'activation couramment utilisée est la fonction sigmoïde, qui est définie comme suit :

$$VAj = f(VEj) = 1 / (1 + \exp(-VEj))$$

Cette opération est répétée jusqu'à ce que les valeurs d'activation des neurones de la couche de sortie soient calculées. La différence entre ces valeurs et les valeurs désirées représente l'erreur d'apprentissage, appelée delta (Δ).

Ensuite, dans l'étape de rétropropagation, on distribue cette erreur à travers les couches du réseau, en remontant des sorties vers les entrées (en arrière), afin d'ajuster les poids du réseau

dans la prochaine phase. On calcule l'erreur (Δ) de chaque neurone s de la couche de sortie en utilisant la formule suivante :

$$\Delta s = VA\ s * (1 - VA\ s) * (\text{valeur désirée de S} - VA\ s)$$

Ensuite, on calcule l'erreur de chaque neurone caché j en utilisant la formule suivante :

$$\Delta j = VAj * (1 - VAj) * \Sigma k \in succ(j)\ wjk * \Delta k$$

Dans ces formules, Wij représente le poids entre le neurone i et le neurone j, VAi est la valeur d'activation du neurone i, wjk est le poids entre le neurone j et le neurone k, et Δk est l'erreur du neurone k.

Ces calculs d'erreurs sont utilisés pour ajuster les poids du réseau dans la phase suivante de l'apprentissage, afin de minimiser l'erreur globale du réseau et améliorer ses performances.

• **Mise à jours des poids :** Lorsque l'erreur d'apprentissage est distribuée sur tous les neurones de la couche cachée, on met à jour les poids et les biais en utilisant les formules suivantes :

$$wij = wij + \alpha * VAi * \Delta j \quad bj = bj + \alpha * \Delta j$$

Où α est le taux d'apprentissage choisi par l'utilisateur, qui est un nombre compris entre 0 et 1. Ce taux détermine l'ampleur des ajustements effectués lors de la mise à jour des poids et des biais.

Ce processus de mise à jour des poids et des biais est répété autant de fois qu'il y a d'exemples dans la base d'apprentissage. Une fois cette étape terminée, on calcule la somme des erreurs au carré (SEC) à l'aide de la formule suivante :

$$SEC = (1/n) * \Sigma n\ i=1\ (y\_desk - yk)^2$$

Où y_desk est la valeur désirée dans l'exemple d'apprentissage et yk est la valeur de sortie calculée par le MLP pour cet exemple.

L'objectif est de minimiser la valeur de SEC en ajustant les valeurs de pondération des nœuds et des liaisons (c'est-à-dire les poids et les biais) du réseau. Si la valeur de SEC n'est pas optimale, on peut répéter la procédure avec de nouvelles valeurs initiales des poids (wij) et des biais (bj) afin de chercher une meilleure configuration pour le réseau.

La phase d'apprentissage d'un réseau de neurone peut donc être résumée par l'algorithme suivant :

Initialisation des poids et biais avec des valeur aléatoire comprises dans intervalle choisi.

Répéter

Pour chaque exemple d'apprentissage faire

Propagation de l'entrée vers l'avant.

Propagation de l'erreur vers l'arrière.

Mise à jour des poids et biais.

Fin pour

Calcul de l'erreur totale SEC.

Jusqu'à(l'erreur totale devient inférieure au SEUIL).

Figure 2.12 -L'algorithme de rétro propagation de gradient**.**

c) Validation :

La phase de validation dans un réseau de neurones intervient après la phase d'apprentissage. Elle utilise la matrice des poids optimaux obtenus pendant l'apprentissage ainsi que les vecteurs d'entrée correspondant aux exemples de la base de test. Pendant la validation, une opération de propagation d'états est effectuée, où les valeurs d'entrée sont propagées à travers le réseau pour générer des valeurs de sortie. La phase de validation permet d'évaluer les performances du réseau sur des données indépendantes et de vérifier sa capacité à généraliser les connaissances apprises pendant l'apprentissage.

11-Domaines d'application réseaux de neurones artificiels[34] :

Traitement d'images : Reconnaissance de caractères et de signatures, compression d'images, reconnaissance de forme, cryptage, classification, etc.

Traitement du signal : filtrage, classification, identification de source, traitement de la parole…etc.

Contrôle : commande de processus, diagnostic, contrôle qualité, asservissement des robots, systèmes de guidage automatique des automobiles et des avions…etc.

Défense : guidage des missiles, suivi de cible, reconnaissance du visage, radar, sonar, lidar, compression de données, suppression du bruit…etc.

Optimisation : planification, allocation de ressource, gestion et finances, etc.

Simulation : simulation du vol, simulation de boîte noire, prévision météorologique, recopie de modèle

12-Les avantages et les limites des réseaux de neurones :

✓ Les avantages[19] :

- **L'apprentissage :** la possibilité d'apprendre et de généraliser les connaissances acquises. ·
- **Le parallélisme massif** : l'architecture d'un réseau de neurone permet le traitement parallèle et rapide des informations. ·
- Une tolérance à l'incertitude très élevée. ·
- Etant une multiple copie d'unités simples (les neurones), ils sont donc facilement extensibles.
- Une facilité d'utilisation car ne nécessitant pas une compréhension approfondie. ·
- Un choix de types, d'architecture et de fonction d'activation de réseaux diverses.

✓ Les limites :

- **L'architecture et le paramétrage du réseau** : Il est difficile de définir pour un problème donné l'architecture et le type adéquat du réseau à utiliser en plus des bons paramètres de réglage. ·
- **Initialisation et codage** : L'état initial du réseau est très important car il a une très grande influence sur la fiabilité et le temps de calcul et de réponse du réseau, donc
- Un mauvais choix des poids initiaux peut conduire à un blocage d'apprentissage et de convergence du réseau vers la solution désirée. ·
- **Boite noire** : Les Réseaux de Neurones Artificiels sont des boites noires, dans la mesure où l'on ne peut connaître et interpréter directement les connaissances acquises par le réseau, donc ces connaissances sont inintelligibles pour l'utilisateur ou l'expert. ·
- Une facilité d'application donnant lieu à de nombreuses implémentations et des choix pas toujours justifiés.

Conclusion

Dans ce deuxième chapitre, nous avons présenté en premier lieu une introduction au domaine de classification automatique des données, où nous avons abordé les notions de base de ce domaine, les différentes catégories de la classification ainsi que les différentes techniques de classification de chaque catégorie .

Dans la deuxième partie, nous avons introduit les définitions essentielles relatives aux réseaux de neurones, et cela en spécifiant leur principe de fonctionnement, leurs types d'apprentissage et leur architecture avec une brève présentation de quelques modèles de ces derniers, où nous avons concentré sur les réseaux de neurones multicouches(MLP) qui sont les plus adaptés à la classification du trafic réseau,et enfin nous avons terminé par les avantages et les limites de cette technique .

Introduction

Les problèmes d'optimisation occupent actuellement une place importante dans la communauté scientifique car il est nécessaire de trouver des solutions optimales à des problèmes très complexes et difficiles à résoudre, où l'on trouve un intérêt croissant d'utiliser la biologie comme source d'inspiration pour résoudre différents problèmes.Ce domaine de recherche se base principalement sur l'extraction des métaphores utiles à partir des systèmes biologiques afin de créer des solutions informatiques efficaces aux problèmes complexes. Les développements les plus appréciables ont été les réseaux de neurones inspirés par le fonctionnement du cerveau, et les algorithmes évolutionnaires.

Cependant, plus récemment, un intérêt croissant pour l'utilisation d'un autre système biologique qui est le système immunitaire qui regroupe un ensemble d'algorithmes capables de résoudre des problèmes assez complexes en s'appuyant sur des comparaisons avec le système immunitaire humain qui doté par des capacité de traitement , l'apprentissage, mémorisation. Pour ces dernières et d'autres raisons et d'utiliser comme une métaphore d'inspiration dans le calcul. Ce domaine de recherche est appelé système immunitaire artificiel.

Dans ce chapitre nous avons introduit le concept d'optimisation combinatoire et ses différentes techniques, en nous concentrant sur système immunitaire artificiel car c'est la méthode l'optimisation que nous allons utiliser pour créer notre modèle de détection d'intrusion.

I-Optimisation Combinatoire

1-Définition :

L'optimisation combinatoire est une branche de l'optimisation mathématique qui a des applications dans l'intelligence artificielle, l'apprentissage automatique, le génie logiciel, l'informatique théorique, les mathématiques appliquées et de nombreux domaines différents. Il est lié aux théories de la complexité computationnelle, aux algorithmes et à la recherche opérationnelle[35].

L'optimisation combinatoire concerne principalement les techniques utilisées pour résoudre les problèmes d'optimisation ; il n'inclut généralement pas d'instructions sur la façon de traduire des problèmes concrets en problèmes mathématiques abstraits correspondants, ou l'inverse. L'optimisation combinatoire est une technique populaire de nos jours pour étudier les algorithmes, notamment ceux utilisés dans la recherche opérationnelle, l'apprentissage automatique et l'intelligence artificielle[35].

2.Classification des méthodes d'optimisation combinatoire

Les techniques présentées dans la figure suivante sont utilisées pour résoudre un problème d'optimisation combinatoire :

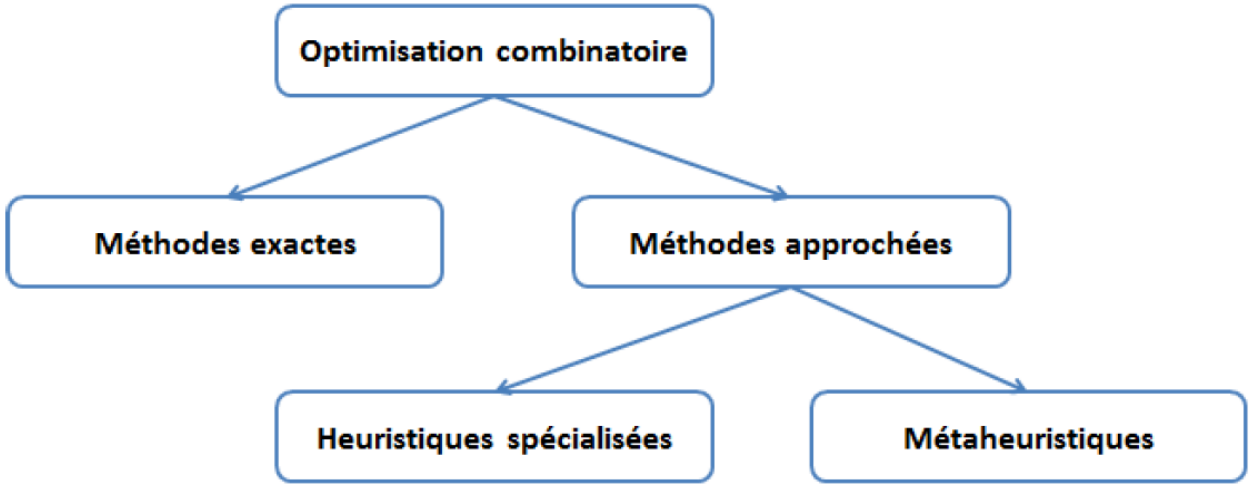


Figure 3.1 – Classification des méthodes d'optimisation.

Méthodes exactes

L'idée de base derrière la méthode exacte est de lister toutes les réponses possibles dans le champ de recherche, souvent implicitement. Cette stratégie comprend des approches pour identifier les défaillances le plus tôt possible (calcul de limites) et des heuristiques particulières pour guider diverses décisions afin d'améliorer l'énumération des solutions.

On retrouve la plupart des méthodes classiques (créées au cours des 30 dernières années) au sein des micro-méthodes, y compris les techniques de programmation dynamique, de classe et d'évaluation (branch and join). Pour des problèmes de taille raisonnable, des méthodes précises peuvent être utilisées pour déterminer la meilleure solution.[36]

a) **Méthodes approchées**

On peut utiliser des méthodes approchées tout en se contentant de chercher une bonne solution lorsque l'on dispose d'un temps de calcul limité, lorsque l'on est confronté à des problèmes difficiles, ou lorsque son problème est volumineux ou compliqué. Dans cette situation, choisir entre une heuristique spécialisée et uneméta heuristique est parfois une option :

- **Heuristique :** Le mot «heuristique» dérive du grec «eurisko», qui signifie «je trouve», donnant lieu au moment bien connu d'Eureka d'Archimède. Les heuristiques, également connues sous le nom de méthodes d'approximation, sont des algorithmes qui offrent

rapidement (en temps polynomial) une solution réalisable, mais pas toujours idéale, à un problème d'optimisation difficile. Une technique heuristique repose souvent sur sa propre structure et est créée pour une situation spécifique.[36]

- **Méta heuristique :**

Les termes grecs méta, qui signifie « au-delà », et heuristique, sont combinés pour former le mot méta heuristique. En fait, ces algorithmes sont conçus pour être des solutions générales qui peuvent résoudre une variété de problèmes sans modifier de manière significative la méthodologie sous-jacente.[36]

I I– System immunitaire

1-Introduction :

Toutes les créatures vivantes sont dotées par un système immunitaire, qui spécialisé dans la défense du corps contre les agents étrangers par exemple Le système immunitaire biologique c'est une façon de se défendre contre les intrus dans un corps particulier.

Pour ce faire, il existe plusieurs cellules qui contribuent à éliminer ces intrus nommé antigènes. Ces cellules participent pour ce qu'on appelle une "Réponse Immunitaire Biologique. Pour que le système immunitaire défende l'organisme contre ces agents ,il doit être capable de faire la distinction entre ce qui appartient au corps(subjectivement)et ce qui n'appartient au corps (non-soi ou étranger).

Le système immunitaire naturel est assez compliqué pour qu'une simulation artificielle soit réalisée d'une façon complète. Par contre, les chercheurs ont réussi à simuler les fonctions les plus pertinentes dans un système immunitaire biologique pour que l'artificiel hérite le maximum des fonctionnalités naturelles dans le domaine de la reconnaissance des formes.

1- Système immunitaire biologique (SIB)

1.1 Définition :

Le système immunitaire biologique est le mieux défini par son rôle, lequel consiste à protéger et à défendre l'organisme hôte contre les éléments qui lui sont nuisibles[37].

L'immunologie est le domaine de la science qui étudie le système immunitaire biologique par l'investigation des mécanismes de défense de l'organisme contre les substances[37] .

L'un des rôles les plus importants d'un système immunitaire biologique (SIB) consiste à protéger l'organisme hôte contre les agressions d'éléments pathogènes, aussi bien internes (provenant du milieu intérieur), tels que par exemple les tumeurs et les cellules cancéreuses, qu'externes (provenant de l'environnement extérieur), tels que par exemple les bactéries, les virus et les parasites. S'ils ne sont pas stoppés et/ou éliminés, ces pathogènes prolifèrent dans l'organisme et génèrent des maladies nuisibles à son bon fonctionnement.

Le système immunitaire est alors confronté aux problèmes de détection, d'identification et de réponse aux pathogènes. Suite à la détection de la présence d'éléments nuisibles dans l'organisme, l'étape d'identification ou reconnaissance consiste à orienter la réaction immunitaire vers les mécanismes les plus appropriés pour déclencher une réponse immunitaire et ainsi éliminer ou stopper les effets indésirables du pathogène[38].

Figure 3.2-Système immunitaire.

2-L'architecteur du system immunitaire

Le corps humain est équipé de plusieurs mécanismes de défense pour éliminer l'intrus.

D'autre part, le système immunitaire a une structure multicouche divisée en deux couches interconnectées, à savoir le système immunitaire inné et le système immunitaire adaptatif ou acquis , chacun ayant une fonction et un rôle différent (Fig. 3.3).

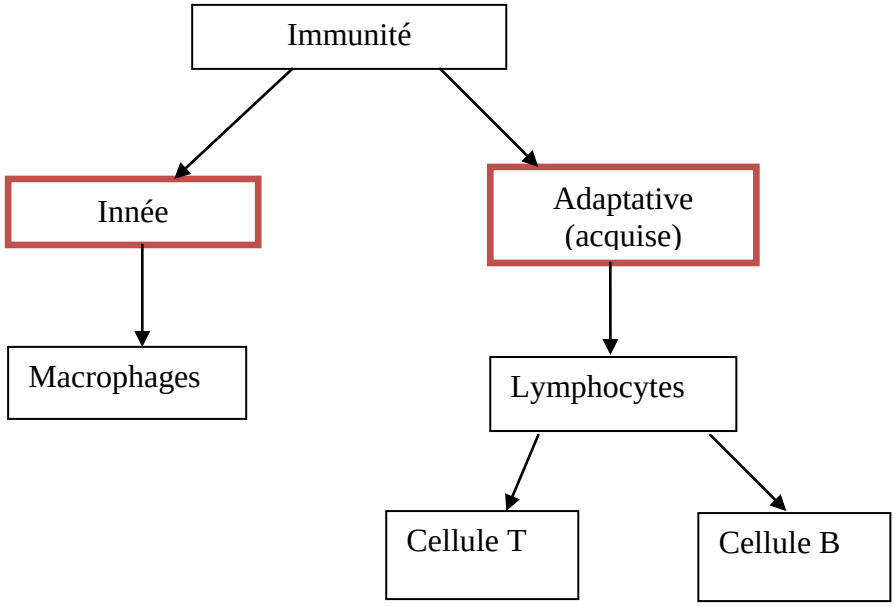


Figure 3.3-Architecture du système immunitaire.

- **L'immunité adaptatif :**La composante adaptative s'organise autour de deux classes cellulaires spécialisées, les lymphocytes T et les lymphocytes B. très vaste et extrêmement diversifié. La taille et la diversité de ce répertoire augmentent la probabilité qu'un lymphocyte individuel rencontre un antigène qui se lie à son récepteur, déclenchant ainsi l'activation et la prolifération cellulaire. Ce processus, appelé sélection clonale, explique la plupart des propriétés de base du système immunitaire adaptatif[39]**.**

✓ **Cellule T** :Un type de globule blanc, cellule T qui fait partie du système immunitaire et provient de la glande thymus. Aide à protéger le corps contre les infections. Ils sont également appelés lymphocytes T et cellules du thymus.

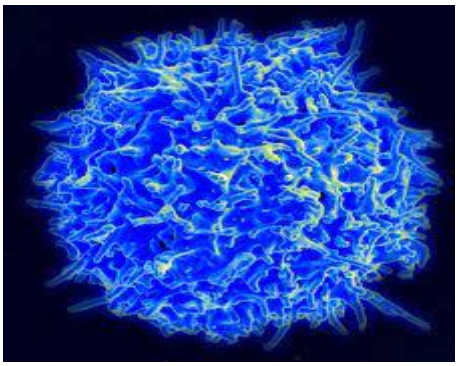


Figure 3.4- Cellule lymphocyte T.

✓ **Cellule B :**Type de globule blanc qui produit des anticorps, les lymphocytes B font partie du système immunitaire et se développent à partir des cellules souches de la moelle osseuse. Aussi appelé lymphocyte B.

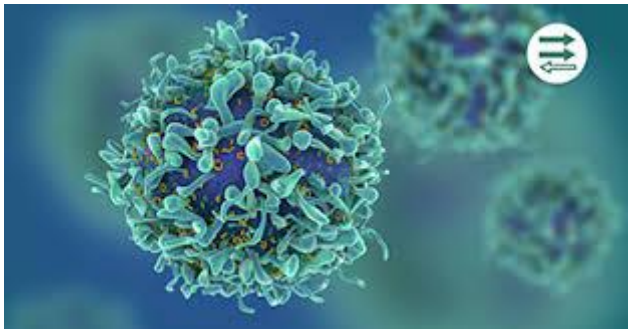


Figure 3.5-Cellule lymphocyte B.

▪ **L'immunité innée** :les mécanismes effecteurs de l'immunité innée, qui comprennent les peptides antimicrobiens, les phagocytes et la voie alternative du complément, sont activés immédiatement après l'infection et contrôlent rapidement la réplication du pathogène infectant. Pour cette raison, contenir l'infection jusqu'à ce que les lymphocytes puissent commencer à y faire face a longtemps été considéré comme la fonction principale de l'immunité innée. Cependant, il est devenu de plus en plus clair que le système immunitaire inné a un rôle beaucoup plus important et fondamental dans la défense de l'hôte[39].

✓ **Macrophages :**sont des cellules appartenant aux globules blancs Rôle dans le processus de phagocytose, tuant les bactéries et libérant des antigènes.est également une grande cellule macrographie qui extrait le calcium de l'os.

Figure 3.6-Macrophages.

3- Concepts immunologiques

3.1 Les anticorps :

Les anticorps sont des molécules complexes appartenant à la famille des immunoglobulines (ce qui explique que l'abréviation courante d'anticorps soit Ig). Fabriqués par les plasmocytes des lymphocytes B activés, la fonction unique des anticorps est de reconnaître et se fixer de façon spécifique aux antigènes. À travers la reconnaissance et la distinction de motifs moléculaires spécifiques, les anticorps jouent un rôle central dans le système immunitaire. Les antigènes sont diversifiés dans leur structure, forçant le répertoire d'anticorps d'être de grande taille[39].

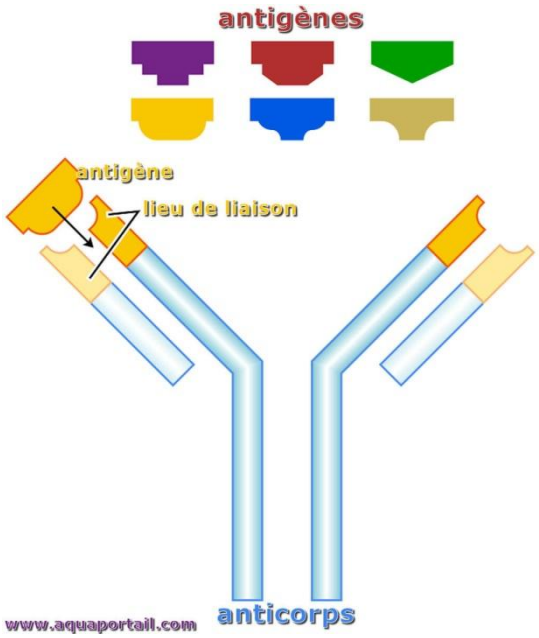


Figure 3.7 -Structure de anticorps .

3.2 Les Antigènes :

Un antigène est une molécule ou un micro-organisme d'origine biologique ou synthétique , classiquement réputé étranger à l'organisme, et capable d'engendrer une réponse immunitaire. Une fois pénétré dans l'organisme, l'antigène n'est pas détecté dans sa totalité par le système immunitaire, ce dernier ne détecte que des déterminants antigéniques, appelés épitropes, sur l'antigène. Ces épitropes se lient de manière complémentaire avec les paratopes des anticorps.

Un même antigène peut comporter plusieurs épitropes (identiques ou différents), ainsi un antigène induit la synthèse de plusieurs différents anticorps (un pour chaque épitope différent). Si ses épitropes sont reconnus comme appartenant au non-soi, alors il est lui-même immédiatement reconnu comme appartenant au non-soi. La reconnaissance épitope/paratopes constitue la base de la réponse immunitaire spécifique permettant la sélection clonale, la sélection des cellules capables de s'attaquer spécifiquement à l'antigène correspondant à un épitrope particulier[40].

4-Comment le système immunitaire assure t-il la protection du corps humain?[39]

Notre corps est protégé par une collection diverse des cellules et des molécules qui collaborent contre n'importe quelle molécule étrangère comme les bactéries ou d'autres envahisseurs. La figure ci-dessous présente une version simplifiée des mécanismes de base de défense immunitaire (figure 3.8), et qui peuvent être résumés par les étapes suivantes :

- Quand un intrus envahit le corps, les cellules de présentation antigénique (APC1) comme les macrophages procèdent à l'ingestion et la digestion de l'antigène rencontré pour le présenter comme des fragments de peptides antigéniques.
- Ces peptides seront liées avec les molécules MHC pour permettre leurs liaisons avec Les cellules T qui ont la capacité de reconnaître la combinaison de peptide / MHC.
- Les cellules T activées par cette identification produisent et sécrètent des lymphokines ou des signaux chimiques pour mobiliser d'autres composants du système immunitaire.
- Les cellules B qui ont aussi des molécules de récepteur complémentaires répondent à Ces signaux. A la différence des récepteurs de cellules T, ceux de cellules B Peuvent reconnaître les parties d'antigènes libres sans les molécules MHC. Après cette activation, les cellules B prolifèrent et se différencient et sécrètent des protéines d'anticorps.
- La liaison entre les anticorps et les antigènes disponibles mènent à la destruction et La suppression des antigènes.
- Un nombre de cellules B et T deviennent des cellules mémoires qui ont une durée de vie illimitée, en permettant l'élimination rapide de l'antigène s'il se présente une autre fois dans l'avenir**.**

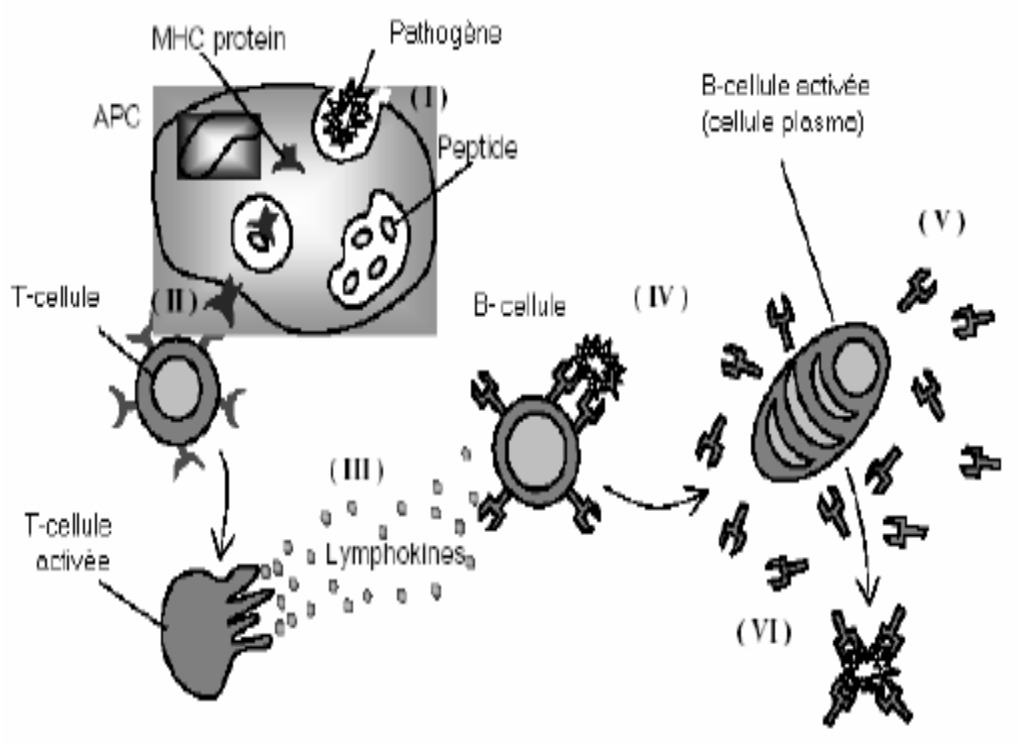


Figure 3.8-La processus de défense immunitaire.

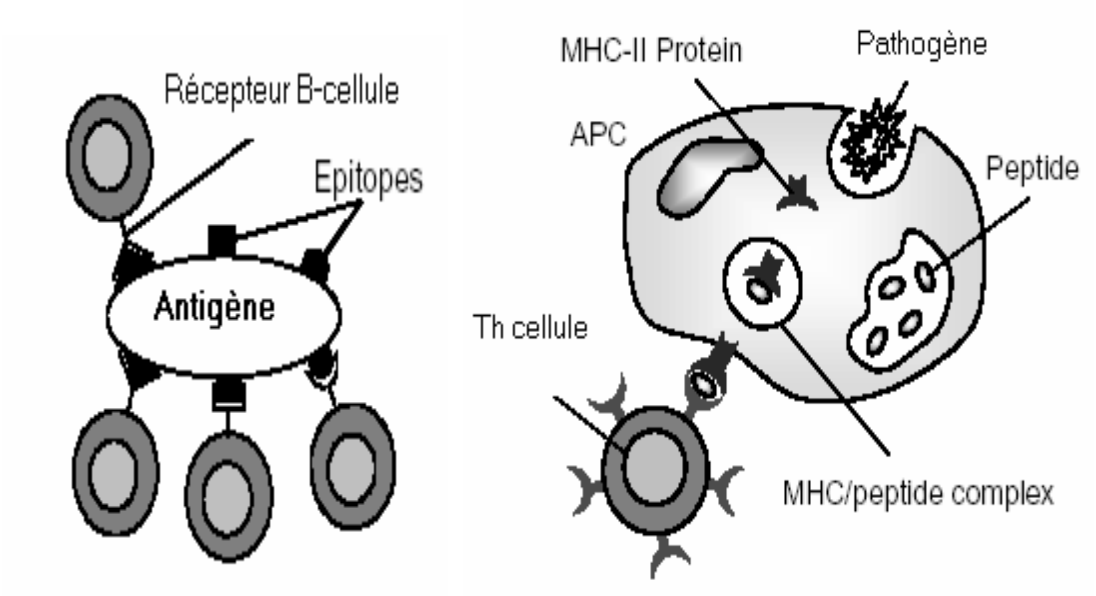
5- Les processus de base d'un système immunitaire [39]

5.1 L'identification dans le système naturel :

Les lymphocytes sont des globules blancs produits dans la moelle osseuse et sont responsables de l'identification pathogènes. Ces lymphocytes ont des récepteurs situés sur leur surface responsables de la reconnaissance des antigènes. Les antigènes et les récepteurs cellulaires doivent avoir des formes complémentaires pour pouvoir se lier ensemble. C'est la liaison du récepteur avec les antigènes qui déclenchent une réponse immunitaire.Les cellules B et T ont une structure semblable mais elles ont une manière de reconnaissancedifférente :

a) Reconnaissance d'un antigène par un récepteur de cellule B.

b) les cellules T ont la possibilité de reconnaître l'antigène qui est présenté par les molécules MHC**.**



(a) **(b)**

Figure 3.9- L'identification dans le système naturel.

5.2 L'activation

La reconnaissance antigène est la première condition préalable à l'activation du système immunitaire pour déclencher une réponse contre le pathogène qui présente l'antigène reconnu.

L'appariement entre un récepteur cellulaire et un l'antigène détermine l'affinité, et cette appariement se produit force proportionnelle à cette affinité. Si l'affinité est supérieure à un seuil donné, appelé "seuil d'affinité", alors le système immunitaire est activé. La nature de l'antigène, le type de cellule reconnaissante et le site de reconnaissance influencent également l'issue d'une rencontre entre un antigène et un récepteur cellulaire.

Le système immunitaire humain contient un organe appelé thymus, qui joue un rôle dans la maturation des cellules T. Après cellules T Sont générés, ils migrent dans le thymus où ils

mûrissent. Pendant cette maturation, toutes les cellules T qui reconnaissent les auto-antigènes sont exclues et cette processus appelé sélection négative. Si une cellule B rencontre un antigène non-soi avec une affinité suffisante, elle prolifère et se différencie en cellules mémoire et effectrices et cette processus nommé sélection clonale. En revanche, si une cellule B reconnaît un auto-antigène, elle pourrait entraîner la suppression, comme proposé par la théorie du réseau immunitaire.

6-Le système immunitaire artificiel :

Dans le but de résoudre des problèmes complexes, des idées inspirées à partir de mécanismes naturels ont été exploitées pour développer des heuristiques inspirées de la nature, le système immunitaire artificiel (SIA) est un paradigme récent qui tente de capturer des caractéristiques intéressantes des systèmes immunitaires naturels (SIN), comme la mémorisation, la reconnaissance de formes, l'apprentissage, et les capacités d'adaptation , la détection d'intrusion dans les réseaux informatiques , la robotique , l'apprentissage machine , etc[40].

6.1 définition :

- C'est un type d'algorithme inspiré du système immunitaire naturel à travers les principes et les caractéristiques du système immunitaire en relation avec l'apprentissage et la mémoire comme outil de résolution de problèmes.

- Selon Dasgupta : "Le système immunitaire artificiel est la composition de méthodologies intelligentes inspirées par le système immunitaire naturel afin de résoudre des problèmes du monde réel" [40].

6.2 Structure de conception d'un system immunitaire artificiel :

Dans une tentative de création d'un modèle commun pour les SIA, les auteurs de [dCT02a] ont adopté un schéma de structure pour la conception d'un algorithme de SIA, nécessitant au moins les éléments de base suivants[38] :

- Une représentation des composants du système (modèles abstraits des cellules immunitaires) ;
- Un ensemble de fonctions pour évaluer l'affinité (la similarité) entre les composants du systèmes ;
- Un ensemble d'algorithmes pour contrôler l'évolution et la dynamique du système immunitaire artificiel.

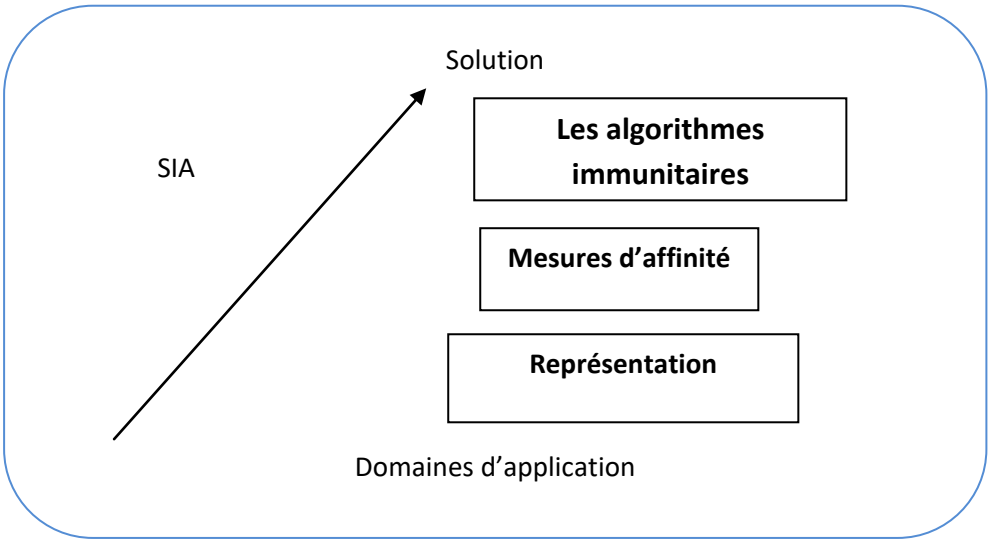


Figure 3.10-Structure de conception d'un system immunitaire artificiel.

6.2.1 La représentation :

Afin de construire un système tel qu'un SIA, un domaine d'application ou une fonction cible sont généralement exigés. A partir de cette base, la façon dont les éléments du système (cellules) seront représentés est considérée. Cette façon est appelée espace de forme (shape space). Il existe plusieurs types d'espaces de forme, tels que Hamming, les valeurs réelles, etc. chacun porte son propre biais et doit être choisi avec précaution [38].

6.2.2 Mesures d'affinité

L'affinité entre un anticorps et un antigène est relative à leur distance . Elle peut être estimée via n'importe quelle mesure de distance entre deux chaînes (ou vecteurs) par exemple par l'utilisation de la distance **Euclidienne**, la distance de **Manhattan** ou la distance de **Hamming**[18]. Si on considère un anticorps Ab = <Ab1, Ab2, …AbL> et un antigène Ag = <Ag1, Ag2, …AgL>, alors la distance D peut être calculée selon l'une des distances précédentes qui seront présentées respectivement dans la figure suivante[42]:
la distance de Manhattan :

$$D=\sum_{i=1}^{L}|Abi-Agi|$$

la distance de Hamming :

$$D=\sum_{i=1}^{L}\delta i \; ou \; \delta \begin{cases} 1 \; si & Abi \neq Agi \\ 0 \; sinon \end{cases}$$

6.2.3 Les algorithmes immunitaires

L'algorithme immunitaire artificiel (AIA) est une méthode de calcul qui s'inspire du système immunitaire naturel des organismes vivants. L'AIA est utilisé pour les tâches d'optimisation, d'apprentissage et d'analyse de données, il existe plusieurs types:

6.2.3.1 La sélection négative :

Le but de la sélection négative est de fournir une tolérance aux cellules du soi. Il traite de la capacité du système immunitaire à détecter des antigènes inconnus sans réagir aux cellules du soi. Lors de la génération des lymphocytes T, les récepteurs sont fabriqués par un processus de réarrangement génétique pseudo-aléatoire. Ensuite, ils subissent un processus de censure dans le thymus, appelé sélection négative.

Les lymphocytes T qui réagissent contre les auto-protéines sont détruits ; ainsi, seuls ceux qui ne se lient pas aux protéines du soi sont autorisés à quitter le thymus. Ces cellules T matures circulent ensuite dans tout le corps pour remplir des fonctions immunologiques et protéger le corps contre les antigènes étrangers[41].

L'algorithme de sélection négative est l'un des modèles computationnels de discrimination soi/non-soi, d'abord conçu comme une méthode de détection de changement. C'est l'un des les premiers algorithmes AIS qui ont été appliqués dans divers monde réel applications. Depuis sa conception, il a attiré de nombreux Chercheurs et praticiens de l'AIS et a traversé quelques évolution phénoménale. Malgré l'évolution et la diversification de cette méthode, les principales caractéristiques d'une sélection négative algorithme décrit par Forrest et al [30].

✓ **Algorithme de selection negative :**

L'algorithme de sélection négative est basé sur la génération d'un ensemble de détecteurs qui sera chargé de détecter les éléments du non-soi.

L'algorithme de la sélection négative déroulera comme suit : Étant donné l'ensemble des modèles de soi à être protégé (P), générer un ensemble (M) de détecteurs qui n'identifie aucun élément appartenant à l'ensemble P. Le processus itératif pour produire l'ensemble des détecteurs (M) est décrit comme suit (Figure 3.11 (a)) :

3. Générer des éléments candidats (C) de type chaîne d'une façon aléatoire.

4. Déterminer l'affinité entre chaque élément en (C) avec tous les éléments de l'ensemble de soi (P).

5. **Si** l'affinité d'une chaîne dans (C) avec au moins une chaîne dans (P) est plus grande ou égale à un seuil d'affinité prédéfini Alors Cette chaîne reconnaît l'ensemble de soi, ce qui implique qu'elle doit être éliminée.

Sinon la chaîne est ajoutée à l'ensemble de détecteurs (M).

Une fois que l'ensemble de détecteurs est produit, l'étape suivante de l'algorithme consiste à contrôler le système contre la présence des modèles de non soi tel que chaque élément détecté par les détecteurs générés est considéré comme un élément de non soi[42].

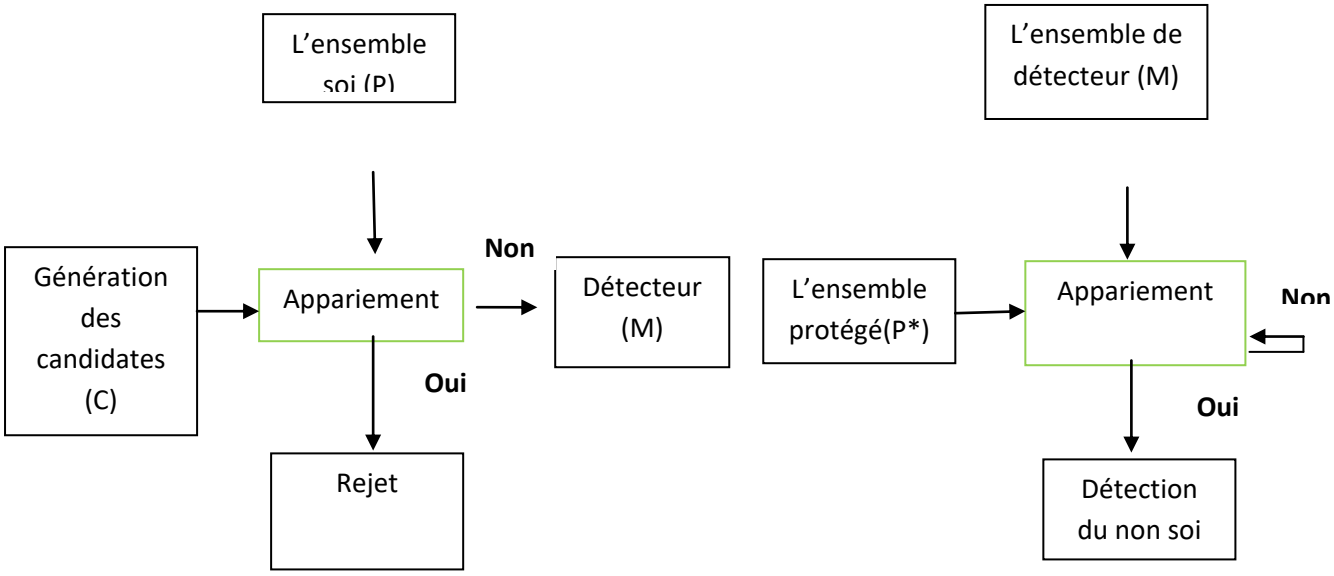


Figure 3.11-La structure général de l'algorithme se sélection négative.

6.2.3.2 L'algorithme de sélection clonale :

L'algorithme de sélection clonale s'inspire de la théorie de la sélection clonale de l'immunité acquise. La théorie de la sélection clonale a été proposée pour rendre compte du comportement et des capacités des anticorps dans le système immunitaire acquis. S'inspirant des principes de la théorie de l'évolution de la sélection naturelle darwinienne, la théorie propose que les antigènes sélectionnent les lymphocytes (lymphocytes B et T). Lorsqu'un lymphocyte est sélectionné et se lie à un déterminant antigénique, la cellule prolifère en faisant des milliers de copies d'elle-même et se différencie en deux types de cellules (cellules à plasma et à mémoire).

Les cellules plasmatiques ont une courte durée de vie et produisent de grandes quantités de molécules d'anticorps, tandis que les cellules à mémoire vivent longtemps chez l'hôte, anticipant la reconnaissance future du même déterminant (voir chapitre 2). La théorie suggère que, à partir d'un répertoire initial de cellules immunitaires générales, le système est capable de se modifier lui-même (composition et densité des cellules et de leurs récepteurs) en réponse à l'expérience de l'environnement. Grâce à un processus aveugle de sélection et à la variation accumulée à grande échelle de plusieurs milliards de cellules, le système immunitaire acquis est capable d'obtenir les informations nécessaires pour protéger l'organisme des dangers pathogènes spécifiques de l'environnement.

Il suggère également que le système doit anticiper (deviner) l'agent pathogène auquel il sera exposé et qu'il doit être exposé à l'agent pathogène pouvant nuire à l'organisme avant

qu'il puisse obtenir les informations nécessaires pour se défendre. Les principes de traitement de l'information de la théorie de la sélection clonale décrivent une stratégie d'apprentissage générale. Cette stratégie implique une population d'unités d'information adaptatives (représentant chacune une solution-problème ou un composant) soumises à un processus de sélection concurrentiel, qui, combiné à la duplication et à la variation résultantes, améliore en définitive l'adaptation des unités d'information à leur environnement.

Le modèle général implique la sélection d'anticorps (solutions candidates) en fonction de l'affinité, soit par correspondance avec un profil antigénique, soit par l'évaluation d'un profil par une fonction de coût. Les anticorps sélectionnés sont soumis à un clonage proportionnel à l'affinité et l'hyper mutation des clones inversement proportionnelle à l'affinité des clones. L'ensemble clonal résultant entre en compétition avec la population d'anticorps existante pour devenir membre de la prochaine génération. De plus, les membres de la population de faible affinité sont remplacés par des anticorps générés de manière aléatoire. La variante de reconnaissance de modèle de l'algorithme comprend la maintenance d'un ensemble de solutions de mémoire qui, dans son ensemble, représente une solution au problème [43].

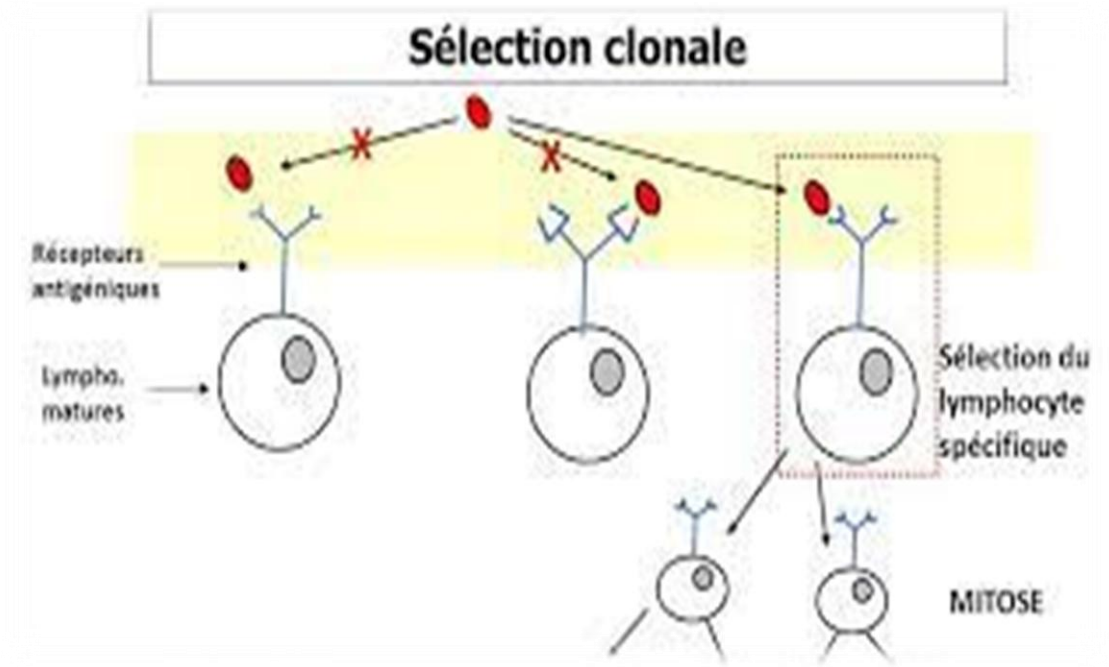


Figure 3.12- Sélectionner les lymphocytes B[44].

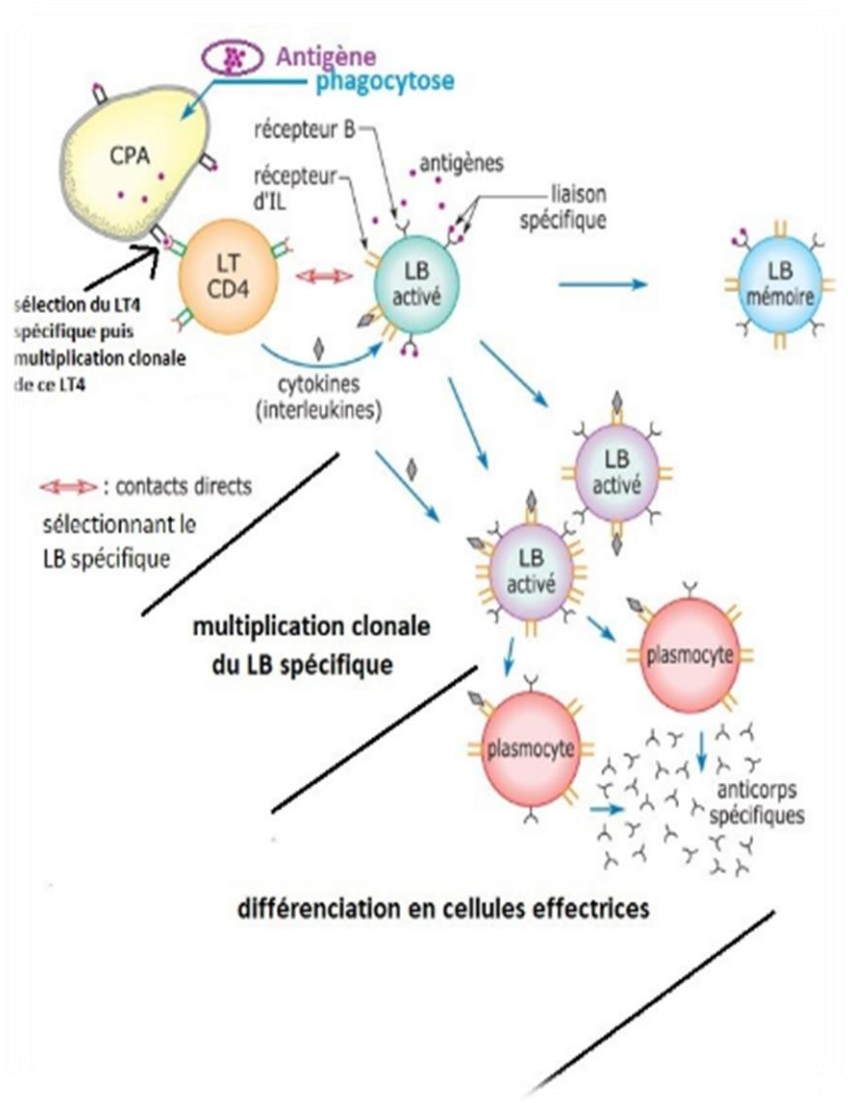


Figure 3.13- Reproduire et apprivoiser les lymphocytes B[43].

Les étapes de base de l'algorithme CLONALE sont résumées comme suit [42]:

1. Générer une population (M) de solution candidate.
2. Déterminer l'affinité de chaque solution candidate avec un ensemble d'antigènes.
3. Choisir les n1 meilleurs éléments de (M) dont ils ont les plus hautes affinités et produire des copies de ces individus proportionnellement à leur affinité avec l'antigène : l'élément qui possède la plus haute affinité aura le plus haut nombre de clones et réciproquement.
4. Muter toutes ces copies avec un taux inversement proportionnel à leur affinité avec l'antigène : l'élément qui possède la plus haute affinité aura un taux de mutation faible et réciproquement.
5. Sélectionner n2 éléments à partir des clones mutés dont ils ont la plus haute affinité pour remplacer les n1 anticorps à l'origine choisis par ces mutants.
6. Remplacer les cellules de faible affinité par des nouvelles cellules aléatoires.
7. Répéter les pas 2 à 5 tant que certains critères sont vérifiés.

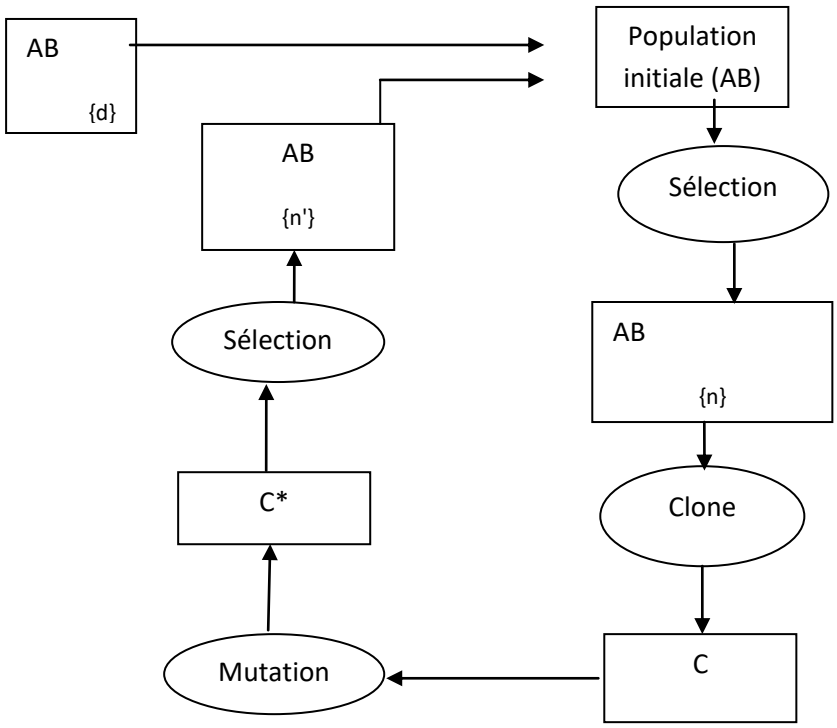


Figure 3.14-Une représentation de l'algorithme de sélection clonale[42].

7- Les domaines d'application d'un system immunitaire [44]:

Comme vu dans les sections précédentes, le système immunitaire artificiel possède une variété de modèles de telle sorte que chaque modèle est basé sur une partie particulière de fonctionnement du système immunitaire humain. Cette diversité permet d'utiliser le système immunitaire artificiel à plusieurs secteurs d'application pour des buts différents. D'une manière générale, parmi ces secteurs on peut citer :

7.1 La sécurité des ordinateurs :

La sécurité des ordinateurs est une application directe de la métaphore du système immunitaire humain. Plusieurs travaux intéressants ont été proposés afin d'exploiter les principes de base de la détection et l'élimination, employés par le système immunitaire humain dans la sécurité des systèmes informatiques. Un travail très intéressant, et qui est considéré parmi les premières tentatives dans ce secteur de recherche et celui de Stéphanie Forrest et son groupe.

Dans ce travail, le problème de la protection des systèmes informatiques est vu comme une instance d'un problème de la discrimination entre le soi et le non soi ce qui signifie la capacité de distinction entre les utilisateurs légitimes et les données non infectées qui constituent le soi et les virus et les utilisateurs non autorisés qui constituent le non soi.

7.2 La détection et l'élimination des virus informatiques :

Okamoto et Ishida ont proposé un système multi agent basé AIS. Ce système de détection de virus opère dans un environnement distribué et hétérogène. L'algorithme de la sélection négative a été utilisé comme une méthode d'authentification de fichier. La détection des virus est réalisée via l'appariement entre les informations propres d'un fichier tel que les premiers bits de l'entête du fichier, sa taille, le chemin d'accès et le fichier de l'hôte. La neutralisation des virus est faite par la réécriture des informations initiales sur le fichier infecté. Le système est composé de quatre types d'agents qui sont :

- Les agents anticorps qui détectent les virus sur les hôtes locaux. - Les agents tueurs qui neutralisent les virus par les réécritures des informations initiales sur les fichiers infectés.

- Les agents de copie qui copient les fichiers non infectés qui sont équivalents aux fichiers infectés à partir des différents hôtes.

- Les agents de contrôle qui aident la communication entre les différents agents.

7.3 Optimisation :

Le problème d'optimisation consiste à trouver l'ensemble absolu des meilleures conditions admissibles pour atteindre un certain objectif. Les problèmes d'optimisation apparaissent dans plusieurs secteurs d'application. Pour cette raison, ce problème est caractérisé par l'existence de plusieurs travaux qui ont exploité le système immunitaire artificiel afin de résoudre les différents problèmes d'optimisation. Par exemple le travail de De Castro & Von Zuben dont le but principal est le développement d'un algorithme approprié pour le problème d'optimisation, la reconnaissance de forme.

Ce travail se focalise sur le principe de la sélection clonale et la maturation d'affinité lors d'une réponse immunitaire adaptative afin de résoudre des problèmes complexes tels que l'optimisation combinatoire et l'optimisation multi modale.

7.4 Robotique :

Plusieurs tentatives ont été faites pour appliquer le principe du réseau immunitaire pour contrôler les grandes populations de robots dont le but principal est d'obtenir les propriétés de base du système immunitaire qui sont l'auto organisation et le comportement de groupe. Le travail de Mitsumoto et al est parmi les premiers travaux, ils ont essayé de créer un groupe de robots qui se comportent d'une façon autonome pour chercher l'alimentation sans aucun mécanisme de contrôle global. L'idée principale dans ce travail est l'interaction entre les robots au niveau local. Les auteurs emploient trois métaphores immunologiques principales.

La première métaphore est les cellules B, où un robot représente une cellule B dont chaque robot possède une stratégie particulière pour trouver l'alimentation. La deuxième est le réseau immunitaire pour garantir l'interaction entre ces robots. La troisième est le calcul de stimulation des cellules B, où le robot qui est le plus stimulé alors sa stratégie est la meilleure pour être prise en considération. Suite à ce travail, plusieurs travaux ont été proposés dans ce domaine de recherche.

Conclusion :

Nous avons pu voir qu'un grand nombre de méthodes d'optimisation existe pour résoudre un problème combinatoire. Dans ce chapitre, nous avons exposé les définitions essentielles à la compréhension de travail du point de vue optimisation et celui du système immunitaire qui sont détaillées dans la deuxième partie de ce chapitre.

Enfin, nous avons présenté deux algorithme qui tomber sur le système immunitaire qui utilisées pour résoudre les problèmes d'optimisation qui sont sélection négative et sélection clonale. Alors, on a donné pour chacune de ces méthodes sa définition, son algorithme Ces méthode seront utilisées dans ce travail pour optimiser modèle MLP dans le but d'augmenter la performance de classification du ce modèle.

Introduction :

Dans ce chapitre, nous abordons l'utilisation du modèle de réseau de neurones artificiels MLP pour la classification supervisée des connexions TCP/IP de la base de données KDD Cup 1999. L'objectif est de développer un système de détection d'intrusions en classifiant le comportement des connexions en attaque ou normal.

Nous décrivons les différentes étapes de notre travail, en commençant par une présentation de la base de données KDD Cup 1999. Nous expliquons ensuite les mesures de prétraitement que nous avons appliquées à cette base de données afin d'améliorer sa qualité et sa pertinence. Ensuite, nous explorons la méthode d'optimisation que nous avons mises en œuvre pour accroître les performances de notre modèle.

Ce chapitre présente donc une approche complète, de la préparation des données à l'optimisation du modèle, pour développer un système de détection d'intrusions basé sur le comportement des connexions TCP/IP.

1-Environnement de programmation:

Dans notre projet, nous avons utilisé les environnements de développement Jupyter notebook, qui est un environnement de développement interactif (IDE) open source basé sur le Web utilisé pour l'analyse de données, le calcul scientifique et l'apprentissage automatique. Il permet aux utilisateurs de créer et de partager des documents contenant du code en direct, des équations, des visualisations et du texte narratif. Jupyter Notebook est construit sur le noyau Python, qui prend en charge l'informatique interactive dans plusieurs langages de programmation, notamment Python, R et Julia, et dans notre projet, nous avons utilisé Python[45].

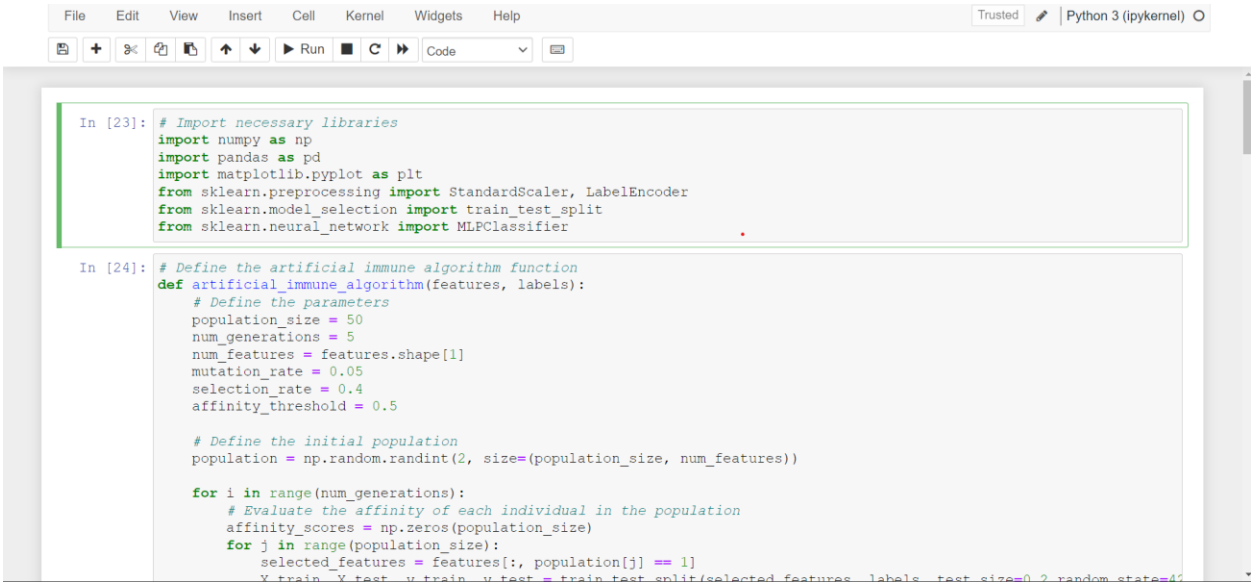


Figure 4.1- jupyter notebook.

Un Jupyter Notebook se compose d'une série de cellules pouvant contenir du code, du texte Markdown ou du texte brut. Les utilisateurs peuvent exécuter des cellules de code en temps réel et voir les résultats immédiatement. La sortie des cellules de code peut être affichée sous forme de texte, de tableaux, de graphiques ou d'autres visualisations. Les cellules Markdown permettent aux utilisateurs de créer du texte formaté à l'aide d'un langage de balisage simple. Les cellules brutes permettent aux utilisateurs d'inclure n'importe quel type de texte sans aucun traitement[45].

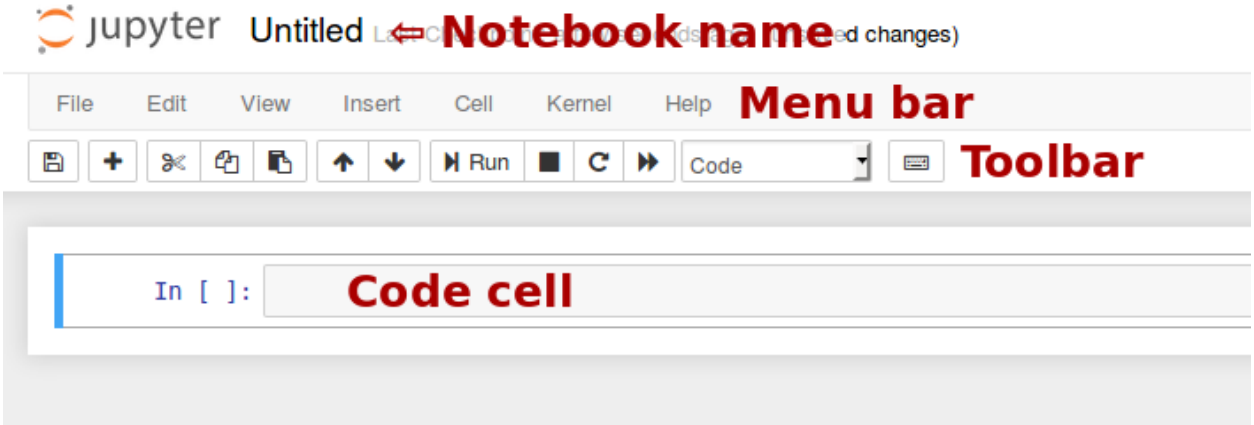


Figure 4.2 - Notebook user interface.

Jupyter Notebook fournit une gamme de fonctionnalités qui en font un outil efficace pour l'analyse de données et le calcul scientifique. Il prend en charge les raccourcis clavier pour les opérations courantes, telles que l'exécution de cellules, l'enregistrement de blocs-notes et la création de nouvelles cellules. Il comprend également une gamme d'extensions qui permettent aux utilisateurs de personnaliser l'interface utilisateur, d'ajouter de nouvelles fonctionnalités et de s'intégrer à d'autres outils[45].

Dans ce travail, où on souhaite élaborer un modèle de détection d’intrusions puissant capable de classifier les connexions TCP/IP en deux catégories : normale ou attaque, on doit passer par les étapes principales que tout modèle de classification doit les suivre. Ces étapes sont résumées dans trois phases principales illustrées dans la figure 4.3 qui sont : le prétraitement, l’apprentissage et enfin la phase de test .

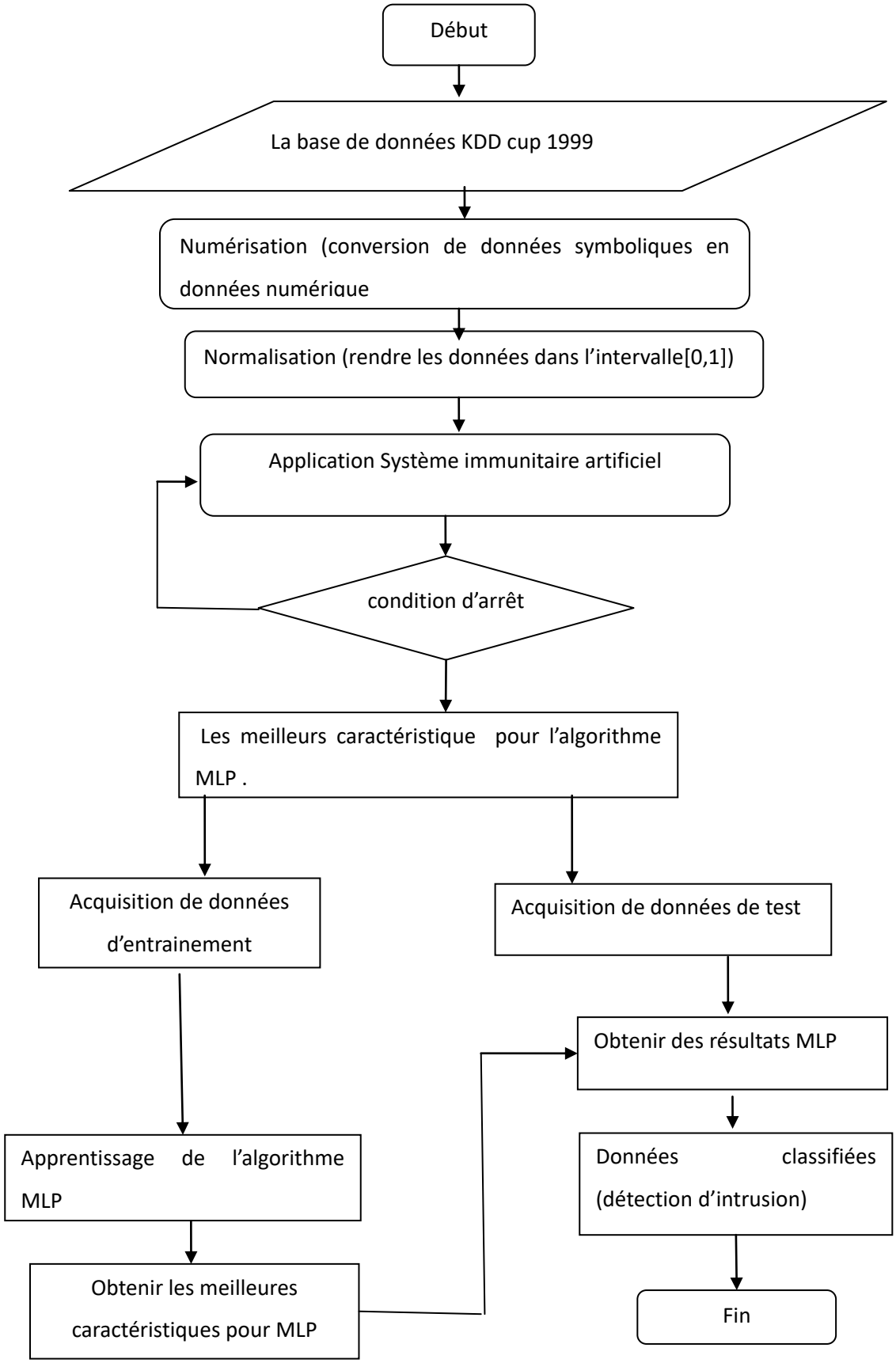


Figure 4.3 -Organigramme de fonctionnement de notre modèle de détection d'intrusion.

2-jeu de données :

Dans ce projet, Nous avons utilisé le jeu de données KDD Cup 1999. est un ensemble de données largement utilisé pour évaluer les systèmes de détection d'intrusion. Il contient des données de trafic réseau collectées à partir d'un environnement de simulation qui émule un réseau local typique de l'US Air Force. L'ensemble de données a été créé dans le cadre du troisième concours international d'outils de découverte de connaissances et d'exploration de données (KDD Cup) qui s'est tenu en 1999.

L'ensemble de données se compose d'un ensemble de caractéristiques de trafic réseau prétraitées, telles que le type de protocole, le type de service, les adresses IP source et de destination et d'autres statistiques réseau. Chaque enregistrement de l'ensemble de données correspond à une connexion réseau et possède une étiquette indiquant s'il s'agit d'une connexion normale ou d'une connexion d'attaque. Il existe 23 types d'attaques différents présents dans l'ensemble de données.

L'ensemble de données KDD Cup 1999 utilisé dans le projet est un fichier CSV.

3-Algorithme système immunitaire artificiel :

L'algorithme immunitaire artificiel (AIA) est une méthode de calcul qui s'inspire du système immunitaire naturel des organismes vivants. L'AIA est utilisé pour les tâches d'optimisation, d'apprentissage et d'analyse de données[46].

Le système immunitaire naturel a la capacité de reconnaître et d'éliminer les envahisseurs étrangers tels que les virus, les bactéries et autres agents pathogènes. Pour ce faire, il utilise des mécanismes complexes de reconnaissance des antigènes, de production d'anticorps et de mémoire immunitaire. L'AIA imite ces mécanismes en utilisant diverses techniques de calcul telles que la théorie des réseaux immunitaires, la théorie de la sélection clonale et la théorie de la sélection négative[46].

Dans ce projet, nous avons utilisé un algorithme d'immunité artificielle pour sélectionner les meilleures caractéristiques pour la détection d'intrusion . La fonction prend les caractéristiques et les étiquettes en entrées et définit les paramètres tels que la taille de la population, le nombre de générations, le taux de mutation, le taux de sélection et le seuil d'affinité. La fonction génère une population initiale de chaînes binaires de 0 et de 1 représentant la présence ou l'absence d'entités. L'affinité de chaque individu dans la population est évaluée à l'aide d'un classificateur de réseau

neuronal entraîné sur les caractéristiques sélectionnées. Les parents sont sélectionnés en fonction de l'opérateur de sélection et la progéniture est générée à l'aide de l'opérateur de mutation. Les clones sont ensuite éliminés à l'aide de l'opérateur de sélection négative, et les parents, la progéniture et les clones sont combinés pour former la nouvelle population. Le meilleur individu de la population finale est renvoyé comme caractéristiques sélectionnées.

4 -Réseaux de neurones multicouches(MLPClassifier):

Dans ce projet, nous avons utilisé un réseau de neurones multicouches (MLP Classifier) pour former notre modèle. Il s'agit d'un réseau neuronal direct composé de plusieurs couches de nœuds (également appelés neurones) qui sont complètement connectés aux nœuds des couches adjacentes. MLPClassifier utilise un algorithme de rétro propagation avec descente de gradient aléatoire pour connaître les poids des connexions entre les nœuds du réseau.

Voici quelque détails de l'architecteur MLP que j'ai utilisé :

1.Configurer la couche masquée :la grillé MLP dans mon code se compose de quatre couche cachées. Chaque couche cachée contient 10 neurones, la configuration spécifique que j'ai utilisé est (10,10,10,10) qui indique quatre couches cachées chacune contenant 10 neurones.

2.Fonction d'activation : vous avez utilisé la fonction d'activation de l'unité linéaire rectifiée(RELU) pour les couches cachées de votre réseaux MLP. RELU est une fonction d'activation largement utilisé qui introduit la non-linéarité du réseaux et aide à capturer des modèles et des relations complexes au sein des données.

3.Couche de sortie :la couche de sortie du réseaux MLP est automatiquement sélectionnée en fonction du nombre de classes dans votre problème de classification .

Etant donné que je n'ai pas explicitement spécifié le nombre de neurones de sortie dans mon code ,MLPClassifier sélectionnera automatiquement le nombre approprié en fonction du nombre d'étiquettes uniques dans votre ensemble de données.

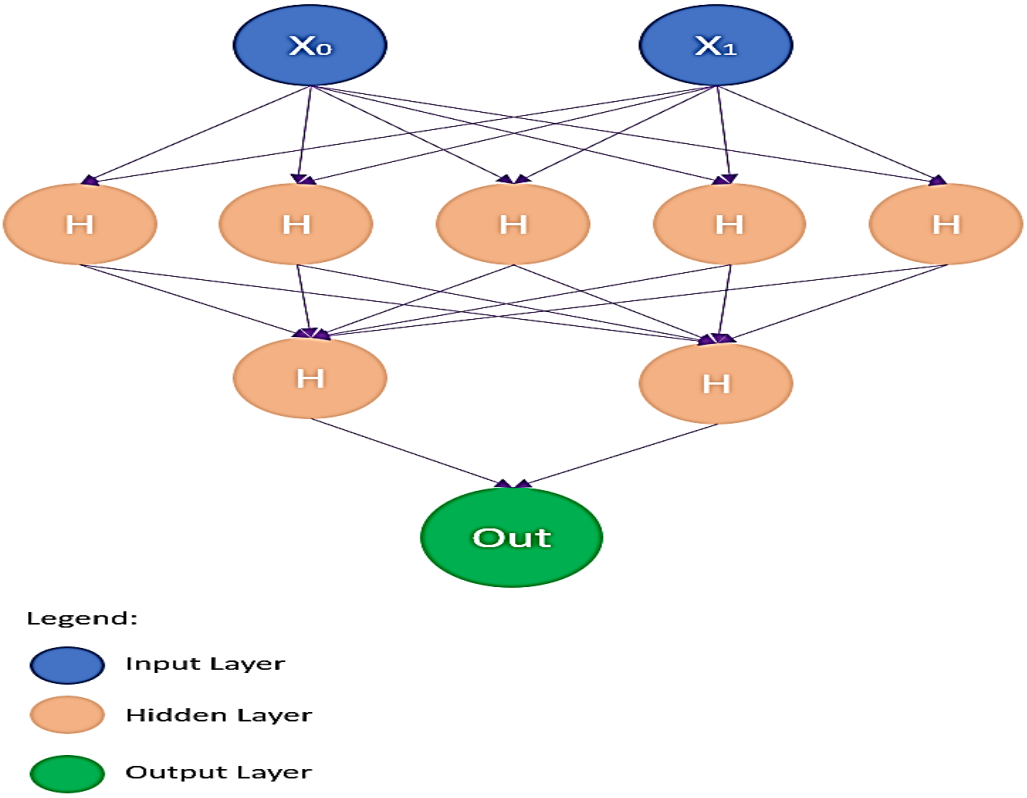


Figure 4.4 - Architecture de réseau neuronal multicouche (MLP Classifier).

La sélection de caractéristiques est effectuée à l'aide de la fonction d'algorithme immunitaire artificiel, et les caractéristiques sélectionnées sont utilisées pour former un classificateur de réseau neuronal. Le classificateur de réseau de neurones est initialisé avec deux couches cachées de 10 neurones chacune et la fonction d'activation Relu. Le taux d'apprentissage est fixé à 0,1. Le réseau de neurones est entraîné sur les fonctionnalités sélectionnées.

5-Traitement du modèle d'apprentissage :

Dans la phase de formation, les données sont divisées en 80% à des fins de formation et 20% à des fins de test.

La fonction d'activation utilisée est ReLU(unité linéaire rectifiée), une fonction d'activation qui définit toute valeur d'entrée négative sur zéro et transmet les valeurs d'entrée positives sans modification.

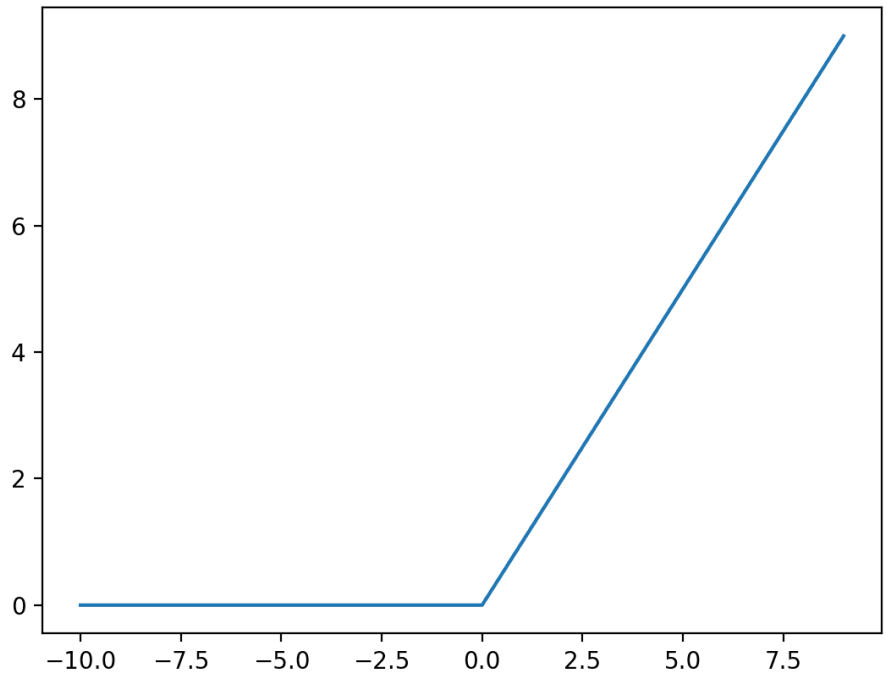


Figure 4.5 - Fonction d'activation Relu.

6-Principe de reconnaissance :

Dans ce projet, un modèle d'apprentissage utilisant des réseaux de neurones a été formé par la fonction d'un algorithme immunologique artificiel, nous avons utilisé un réseau de neurones multicouches (MLP) (MLPClassifier).

Un algorithme immunitaire artificiel sélectionne les meilleures caractéristiques pour la détection d'intrusions. Les caractéristiques définies sont ensuite utilisées pour entraîner le réseau neuronal à l'aide du réseau neuronal multicouche.

6.1 Charger le jeu de données :

Lire l'ensemble de données KDD Cup 1999 à partir d'un fichier Excel (CSV) et le convertir dans un format pouvant être utilisé par le modèle d'apprentissage automatique.

```
In [25]: # Load the dataset
         data = pd.read_csv('dataset/Dataset.csv')
```

Figure 4.6-Charger le jeu de données.

6.2 Prétraiter les données:

Dans l'apprentissage automatique, le prétraitement est le processus de préparation des données pour l'analyse ou la formation. Dans notre projet, l'étape de prétraitement comprend :

1.Encodage des étiquettes catégorielles en valeurs numériques : la classe Label Encoder de la bibliothèque sklearn.preprocessing est utilisée pour encoder les étiquettes dans l'ensemble de données. Cette étape est nécessaire car les algorithmes d'apprentissage automatique ne peuvent fonctionner qu'avec des données numériques, et la plupart des ensembles de données ont des caractéristiques catégorielles qui doivent être encodées.

2.Codage à chaud des caractéristiques catégorielles : la fonction pd.get_dummies() est utilisée pour convertir les caractéristiques catégorielles de l'ensemble de données en caractéristiques binaires à l'aide d'un codage à chaud. Cette étape est effectuée pour s'assurer que le modèle ne suppose aucune relation ordinale entre les catégories.

3.Mise à l'échelle des fonctionnalités : la classe StandardScaler de la bibliothèque sklearn.preprocessing est utilisée pour mettre à l'échelle les fonctionnalités de l'ensemble de données afin d'avoir une moyenne nulle et une variance unitaire. Cette étape est nécessaire car les caractéristiques mesurées à différentes échelles peuvent biaiser l'algorithme d'apprentissage vers les caractéristiques avec des valeurs plus grandes. En mettant à l'échelle les fonctionnalités, nous nous assurons que toutes les fonctionnalités contribuent de manière égale au modèle.

```
In [13]: # Preprocess the data
labels = data['label']
le = LabelEncoder()
labels = le.fit_transform(labels)
features = data.drop(['label', 'service'], axis=1)
features = pd.get_dummies(features)
scaler = StandardScaler()
features = scaler.fit_transform(features)
```

Figure 4.7- Prétraiter les données.

Dans mon code, vous pouvez utiliser l'approche du système immunitaire artificiel pour effectuer la sélection des fonctionnalités. Voici une explication de certains aspects de la modélisation du système immunitaire :

1. Structure individuelle :

Dans mon code, chaque individu est représenté par un vecteur binaire définissant des traits cochés (1) ou indéfinis (0). La taille du vecteur correspond au nombre de caractéristiques disponibles.

2. Changement et choix :

Le changement se produit lorsque de nouveaux individus sont générés dans l'algorithme. Pour chaque nouvelle génération, vous sélectionnez un groupe de parents en fonction de leur affinité (score de performance), puis créez une progéniture en appliquant des opérations de mutation aux parents sélectionnés.

Les parents sont sélectionnés pour la reproduction en fonction de leur affinité, c'est-à-dire que leur performance a été évaluée à l'aide du classificateur MLP. Les parents les plus compatibles (ceux qui ont les scores d'affinité les plus élevés) ont une probabilité plus élevée d'être choisis comme parents.

3. Utiliser le système immunitaire pour extraire des fonctionnalités :

Un système immunitaire artificiel est utilisé pour identifier les caractéristiques les plus uniques en termes de détection d'intrusion. L'idée est que les fonctionnalités qui contribuent le plus aux performances du classifier sont préservées, tandis que les fonctionnalités les moins pertinentes sont ignorées.

L'algorithme évalue l'affinité de chaque individu (ensemble de caractéristiques) à l'aide du classificateur MLP. Les individus ayant une meilleure affinité (performance) sont privilégiés et

ont une probabilité plus élevée de se reproduire et de transmettre leurs caractéristiques aux générations suivantes.

4. Calcul de l'attractivité individuelle :

Dans mon code, l'affinité d'un individu est évaluée à l'aide du score de performance du classificateur MLP lors de la classification des données de test. Le score d'évaluation est une mesure de la précision du modèle prédictif sur les données de test. Une affinité plus élevée indique une meilleure performance du modèle pour détecter les interférences.

5. La différence entre les algorithmes génétiques et le système immunitaire :

Les Algorithmes Génétiques (AG) et le Système Immunitaire Artificiel (SIA) sont deux méthodes d'optimisation inspirées de la biologie.

AG utilise des concepts tels que la reproduction, la mutation et la sélection pour résoudre les problèmes d'optimisation en trouvant les meilleures solutions possibles.

L'SIA s'inspire du système immunitaire biologique et utilise des concepts tels que la reconnaissance, la transcription et la sélection négative pour résoudre les problèmes de détection et de classification des anomalies.

La principale différence est que les AG cherchent à améliorer les solutions globales, tandis que le SIA est principalement utilisé pour détecter des anomalies ou des intrusions dans des ensembles de données.

6. Algorithme de sélection de copie :

Sélection clonale:

L'algorithme de sélection clonale est une étape essentielle du système immunitaire artificiel.

Dans mon code, la sélection clonale se produit lorsque des clones sont créés à partir d'individus choisis comme parents. Les clones sont créés en copiant des individus parentaux et en appliquant des processus de mutation pour introduire la diversité.

La sélection clonale favorise les individus ayant une meilleure affinité (performance) pour augmenter la concentration des caractéristiques les plus pertinentes dans la population et ainsi améliorer la détection des interventions.

Pour montrer l'impact de système immunitaire artificiel sur les performances du réseau de neurone, nous avons effectué deux expérimentations pour les mêmes valeurs des paramètres d'apprentissage.

Pour le premier cas, basons uniquement sur l'opération d'apprentissage pour calculer les valeurs optimales des poids et biais qui seront utilisées pour classifier les connexions dans le système de détection d'intrusions.

Dans le second cas, on utilise système immunitaire artificiel pour trouver la solution optimale qui perfectionne les performances du réseau.

7-Premier cas : sans méthodes d'optimisation :

Nous avons effectué plusieurs tests dans notre en réglant à chaque fois les paramétrées d'apprentissage de réseaux de neurones (MLP)(nombre des couches cachées) dont l'objectif de trouver ces paramètres qui donnent les meilleur résultats en termes du taux de réussite (accuracy).

Dans cette amélioration nous avons définie :

Le nombre d'itération 5 comme critère d'arrêt

Le nombre de couche cachée 4 couche et nous avons changé a chaque fois le nombre de neurones pour donnée meilleur résultats.

Algorithme	**MLP**				
Nombre d'exécution	**01eré**	**02ére**	**03ére**	**04ére**	**05ére**
Nombre des couches	02(5,5)	03(5,5,5)	3(10,5,5)	3(10,10,5)	4(10,10,10,10)
Accuracy(%)	56.97	57.52	67.03	67.73	76,45
Temps des exécution	0 hrs 0 min 24 sec	0 hrs 0 min 27 sec	0 hrs 0 min35 sec	0 hrs 0 min 38 sec	0 hrs 0 min 44 sec

Table 4.1– Evaluation de résultats de classification sans méthodes d'optimisation.

Nous observons que la longueur de la période ,plus le nombre de couche est important ,plus le temps d'entrainement est important et aussi le survenue de surentrainement .

8-Deuxième cas : Avec méthodes d'optimisation (système immunitaire artificiel) : Pour appliquer le système immunitaire artificiel on définie les paramètres qui montré dans le figure(4.8) suivants :

```
In [11]: # Define the artificial immune algorithm function
         def artificial_immune_algorithm(features, labels):
             # Define the parameters
             population_size = 50
             num_generations = 5
             num_features = features.shape[1]
             mutation_rate = 0.05
             selection_rate = 0.4
             affinity_threshold = 0.5
```

Figure 4.8 – Les paramètres pour application système immunitaire artificiel.

8.1 Sélection de caractéristiques à l'aide d'un algorithme immunitaire artificiel:

Dans notre projet, le processus de sélection des caractéristiques est mis en œuvre à l'aide d'un algorithme immunologique artificiel. L'algorithme fonctionne en créant un ensemble de chaînes binaires, où chaque chaîne représente un sous-ensemble des caractéristiques qui seront sélectionnées. L'algorithme évalue ensuite l'affinité de chaque individu dans la population en entraînant un réseau de neurones sur un sous-ensemble sélectionné des caractéristiques et en calculant la précision du modèle sur l'ensemble de validation.

L'algorithme sélectionne ensuite les parents et génère une progéniture à l'aide de l'opérateur de mutation. Les descendants sont combinés avec les parents pour former la nouvelle population. L'algorithme élimine ensuite les clones à l'aide de l'opérateur de sélection négative. Le processus est répété pendant un nombre fixe de générations jusqu'à ce que le meilleur individu soit trouvé.

Enfin, le sous-ensemble de fonctionnalités sélectionné est utilisé pour former un réseau de neurones sur les données. Ce processus permet de réduire la dimensionnalité des données d'entrée et d'améliorer les performances du réseau de neurones.

```
In [14]: # Perform feature selection using artificial immune algorithm
         selected_features = features[:, artificial_immune_algorithm(features, labels) == 1]
```

Figure 4.9- Sélection de caractéristiques à l'aide d'un algorithme immunitaire.

artificiel.

8.2-Entraîner le réseau de neurones sur les fonctionnalités sélectionnées

à l'aide de

8.2 Évaluer les performances du réseau de neurones :

Après avoir effectué la sélection des fonctionnalités à l'aide de l'algorithme immunitaire artificiel, l'étape suivante consiste à former un réseau de neurones sur les fonctionnalités sélectionnées. Cela se fait en divisant l'ensemble de données en ensembles de formation et de test, avec 80 % des données utilisées pour la formation et 20 % utilisées pour les tests. Ensuite, une instance de la classe MLPClassifier est créée avec les hyper paramètres spécifiés, y compris le nombre de couches cachées, la fonction d'activation utilisée dans les couches cachées et le taux d'apprentissage utilisé dans la rétro propagation.

Pendant la formation, les poids du réseau neurons sont ajustés à l'aide de la rétro propagation pour minimiser la fonction de perte, qui mesure la différence entre les valeurs prédites et réelles de la variable cible. Le processus d'apprentissage se poursuit jusqu'à ce que la fonction de perte soit minimisée ou qu'un nombre maximal d'itérations soit atteint.

Une fois le réseau de neurones formé, ses performances sont évaluées sur les données de test à l'aide de la méthode de score de l'objet MLPClassifier, qui calcule la précision du réseau de neurones sur les données de test. La précision est définie comme le pourcentage d'instances correctement classées dans l'ensemble de test.

9-Résultat:

Après avoir sélectionné les meilleures caractéristiques pour l'entraînement par l'algorithme d'immunité artificielle et entraîné le modèle avec les caractéristiques sélectionnées, nous avons obtenu une précision de 97,38 %, ce qui est une bon résultat par rapport au premier modèle MLP . L'image suivante montre le résultat obtenu après l'entraînement de notre modèle.

```
# Evaluate the performance of the neural network
score = classifier.score(X_test, y_test)
print(f'The accuracy of the neural network is: {score}')
```

```
The accuracy of the neural network is: 0.9738245030565564
```

Figure 4.10-La précision du modèle obtenu.

Algorithme	MLP+SIA				
Nombre d'exécution	01ére	02 éme	03éme	04éme	05éme
Nombre des couches	02(5,5)	03(5,5,5)	3(10,5,5)	3(10,10,5)	4(10,10,10,10)
Accuracy(%)	76.18	86.23	88.93	90.97	97.38
Temps d'exécution	1 hrs 57	2 hrs 9 min 7 sec	2 hrs 6 min 43 sec	2 hrs 19 min 7 sec	2 hrs 38 min 15 sec.

Table 4.2 - Evaluation de résultats de classification avec système immunitaire artificiel.

9.1 Analyse et comparaison des résultats :

Au vu des résultats précédents, les performances de notre modèle de détection d'intrusions montrent qu'il peut détecter les intrusions, y compris les nouvelles attaques, avec un taux de réussite allant jusqu'à 97%. D'autre part, cela peut illustrer l'importance d'un système immunitaire artificiel qui peut améliorer les performances du modèle en sélectionnant les meilleures caractéristiques. Cette amélioration est montrée sur le tableau suivant :

Nombre de couches cachés	le taux de Accuracy%			Temps d'exécution		
	Sans SIA	**Avec SIA**	**Différence**	**Sans SIA**	**Avec SIA**	**Différence**
2	56.97	76.18	**+19.21**	0 hrs 0 min 24 sec	1 hr 57 min	1hrs 56 min 36 sec
3	57.52	86.23	**+28.71**	0 hrs 0 min 27 sec	2 hrs 9 min 7 sec	2hrs 9min 20 sec
4	76,45	97,38	**+20.93**	0 hrs 0 min 44 sec	2 hrs 38 min 15 sec.	2hrs 38min29sec

Table 4.3 - Comparaison des performances avec et sans système immunitaire artificiel.

D'après les différentes expériences que nous avons faites, nous remarquons que l'effet de nombre de couches cachées a un rôle efficace dans l'amélioration de ses performances de MLP.

Dans notre cas les meilleures performances de réseaux de neurones optimisé qui contient 4 couches cachées au taux de 97%.

Donc le diagramme montre clairement les améliorations apportées par le système immunitaire artificiel aux réseaux de neurones (MLP).

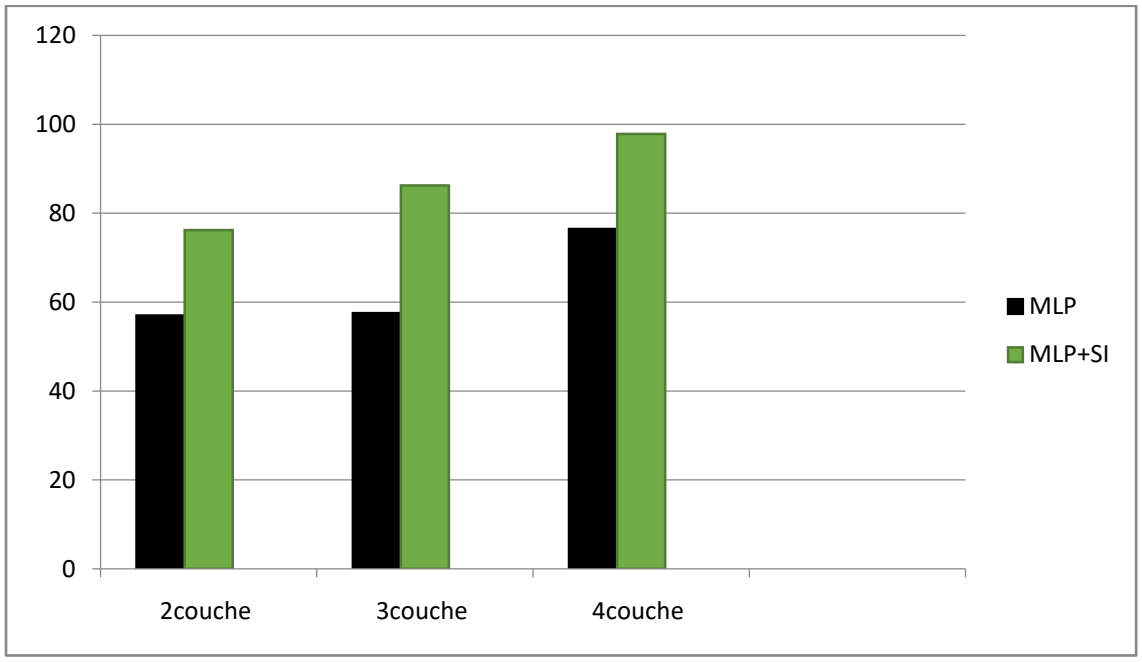


Figure 4.11 -Comparaison des taux de réussite avec des méthodes d'optimisation.

Notion :

Malgré le système immunitaire artificiel a donné des bons résultats pour détection d'intrusion par rapport le perceptron multi couche mais il prend beaucoup de temps pour l'exécution.

Conclusion :

Dans ce dernier chapitre, nous avons présenté les résultats obtenus à partir de plusieurs expérimentations que nous avons effectuées pour arriver au meilleur taux de réussite possible et les améliorations effectuées grâce aux algorithmes de système immunitaire artificiel qui ont été utilisé comme des techniques d'optimisation sur la sélection des caractéristiques.

Les résultats obtenus ont montré l'efficacité des systèmes de détection d'intrusions basés sur les réseaux de neurones, et plus précisément le perceptron multicouche, sur tout lorsqu'il est combiné avec système immunitaire artificiel qui ont montré leur efficacité dans amélioration des performances.

Conclusion générale

L'augmentation des attaques de sécurité et l'espionnage non autorisé ont fait de la sécurité du réseau l'un des principaux sujets à prendre en compte dans l'environnement actuel de communication de données. Par conséquent, les systèmes informatiques deviennent de plus en plus complexes et les approches traditionnelles de la sécurité ne peuvent pas assumer le rôle de protection de manière idéale. Le système de détection d'intrusion est l'un des moyens appropriés pour contrer ces attaques.

Ce qui nous a poussés dans notre mémoire à adopter un modèle de sécurité capable de détecter les intrusions. Pour cela, nous avons utilisé le réseau de neurones dans sa forme générale et la perception multicouche en particulier, car il est le plus approprié pour les données qui ne sont pas linéairement séparables. Il s'agit de classer les connexions TCP/IP en deux catégories : normale ou attaque basée sur les données KDD.CUP 1999.

Nous sommes également concentrés dans notre travail sur le choix du nombre de couches, ce qui à son tour a un effet sur l'amélioration des performances de la perception multicouche, et pour cette raison, nous avons utilisé le système immunitaire artificiel pour l'amélioration en choisissant les meilleures caractéristiques qui nous donnent un bon taux de réussite.

Dans ce travail, nous avons réalisé une étude comparative entre deux modèles de détection d'intrusion l'un se base uniquement sur le réseau de neurone multicouche MLP, et l'autre basé sur la combinaison entre MLP et le système immunitaire artificiel SIA en utilisant ce dernier comme une méthode d'optimisation

Les résultats obtenus montrent l’influence positif des algorithmes d’optimisation sur les performances de réseau de neurones.

Perspective :

Pour ne manipuler plusieurs paramètres et perdre beaucoup de temps dans l’apprentissage et dans le but d’améliorer le SIA, nous proposons d’utiliser une autre méthode d’optimisation e pour choisir les meilleurs paramètres de classification de la base KDD cup 1999.

On propose un système multi-classifier on rajoute une autre méthode de classification à la méthode de sélection clonale d'où chaque méthode est entrainé sur la classe où a prouvé une performance.

Utiliser le système multi agent pour appliquer l’SIA à la détection d’intrusion en temps réel.

A la fin de ce travail, le système immunitaire naturel constitue toujours une source l'inspiration très riche dont le but principal des différentes recherches est la compréhension et l'extraction des mécanismes clefs utilisés par ce système dans l'identification, la détection et l'élimination des intrus afin de construire des systèmes immunitaires pour protéger les systèmes et les réseaux d'une manière efficace.

Bibliographies

[1]: Organisation multi-agent pour la gouvernance de la sécurité informatique avec la norme ISO27001/ISO27002,Ezzrhari, Fatima Ezzahra and Medromi, Hicham, [2012]

[2] :Planification d'une cryptographie des risques dans les systèmes de telecommunication militaires auteur : Bennabti Saida Responsables : Merzougui .R & Ghouali.S Sujet proposé au sein du laboratoire STIC.

[3] : Sécurisation d'un système informatique Soutenu publiquement,[27 / 06 / 2018] (pdf).

[4] :Livre de Sécurité informatique : risques, stratégies et solutions : échec au cyber-roi.

[5] : Détection d'intrusions via des réseaux de neurones optimisés par des méta heuristiques Cherfi Sarra[2020].

[6] : J. Mclean. The speciation and modelling of computer security.Computer, 23(1) [9-16, Janvier 1990]

[7] : https://youtu.be/6SiNFmcD0MM

[8] : Les 10 types de cyber attaques les plus courants Pierre-Louis Lussan Country Manager South-West Europe [Mis à jour : 17 octobre 2022].

[9] : https://youtu.be/Z-TAjREhMeM.

[10]:INTRUSION DETECTIO (Rebecca Gurley Missing)-[2000].

[11] :Importance of Intrusion Detection System (IDS) (January-2011).

[12] : Détection d'intrusions à base des réseaux de neurones et algorithmes génétiques BOUROUH Mouloud KANOUN Zakaria*03 juillet 2017 devant le jury composé de MM.*

[13] :C. Frédéric, "Ids : Intrusion detection systems," http ://www-igm.univ-mlv.fr/ dr/XPOSE2004/IDS/IDSSnort.html,[2005].

[14] :Approche basée sur l'apprentissage profond pour la détection d'intrusion réseau Mr. BABAALI Baligh Mr. LAIB Hamza 2019.

[15] :https://youtu.be/Z6E97TP93NI

[16] :Les systèmes de détection d'intrusion basés sur du machine Learning (prof. Olivier Markevitch & prof. Gianluca Bontempi).

[17]**:**https://math.univ angers.fr/~labatte/enseignement%20UFR/master%20MIM/classificationsupervisee.pdf

[18] :[29/12/2016] . في كشف الاختراق نظمكشفالاختراق , نوارإسماعيل

[19] :Classification supervisée Aperçu de quelques méthodes avec le logiciel R,2012.

[20] :Détection d'intrusions via des réseaux de neurones optimisés par des méta heuristiques(2020).

[21] :NicolasMonmarché ,Algorithmes de fourmis artificielles : applications `a la classification et `a l"optimisation, [2004].

[22] :YOUSSEF FATAICHA, RECHERCHE D'INFORMATION DANS LES IMAGES DE DOCUMENTS, MONTRÉAL, LE 27 DÉCEMBRE [2005].

[23] :Surveillance de procédés à base de méthodes de classification : conception d'un outil d'aide pour la détection et le diagnostic des défaillances.

[24] :Contribution à l'étude du diagnostic des défauts mécaniques par classification non supervisee [2020].

[25] :Classification des images avec les réseaux de neurones Convolutionnels –[2018].

[26] :Détection d'intrusions à base des réseaux de neurones et algorithmes génétiques, BOUROUH .M ,KANOUN .Z ,[16/7/2017].

[27] :https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-rn.pdf.

[28] : Comprendre les réseaux de neurones Mis à jour le30 août 2022.

[29] : LES RESEAUX DE NEURONES ARTIFICIELS, INTRODUCTION AU CONNEXIONNISME : COURS, EXERCICES ET TRAVAUX PRATIQUES. EC2, 1992, Collection de l'EERIE, [J. *Mc Culloch* et W. *Pitts* ,2016].

[30] : : Les réseaux de Neurones formels Et Les systemes Neuro-Flous Pour l'apprentissage par renforcement}, Mezaache, Hatem,2008}, Universite de Batna.

[31] : https://slideplayer.fr/slide/494506/64/video/LES+RESEAUX+DE+NEURONES.mp4.

[32] : Introduction to Machine Learning ,YalınBaştanlar and Mustafa Özuysal,2014

[33] : Chapter Fourteen - Energy-efficient edge based real-time healthcare support system S. Abirami, P. Chitra, in Advances in Computers, [2020]

[34]: Les Réseaux de Neurones Artificiels ,September 2017,Y.Djeriri,U niversity of Sidi-Bel-Abbes.

[35] : Hao, Jin-Kao, Philippe Galinier, and Michel Habib. "Méta heuristiques pour l'optimisation combinatoire et l'affectation sous contraintes." Revue d'intelligence artificielle 13, no. 2 (1999): 283-324.

[36] : Monmarché N. Artificialantbasedalgorithmsapplied to clustering and optimizationproblems (Doctoral dissertation, Ph. D. dissertation, François Rabelais University, Tours, France).

[37]: Etude de la contribution des systèmes immunitaires artificiels au pilotage de systèmes de production en environnement perturbé,[2000].

[38] :Classification du cancer du sein par des approches basées sur les Systèmes Immunitaires Artificiels, Rima Daoudi EpDabladji, 28 septembre 2016.

[39] :Système de détection d'intrusion basé sur un système immunitaire artificiel, TOBBA .A ZEMALI. S 2017.

[40]: Application du system immunitaire artificial pour la reconnaissance des chiffres, Khelil, Hiba and Benyettou, Abd-el-Kader and Belaid, Abdel, Maghreb an Conference on Software Engineering and Artificial Intelligence-MCSEAI'08,2008.

[41]: DipankarDasgupta. “Parallel search for multi-modal function optimization with diversity and learning of immune algorithm”. In : Artificial immune systems and their applications. Springer, 1999, p. 210–220.
[42]:Proposition d’un système immunitaire artificiel pour la détection d’intrusions, LABED .I,2006.
[43] :Conception d’une approche basée agent pour un système immunitaire médical artificiel, SAOULI.S ,03 juillet 2019.
[44] :Application des systèmes immunitaires artificiels à la détection d’intrusion,Slimani.A,2011.

[45] : (Book): Mobini, M., Mobini, Z. and Rabbani, M., 2011. An Artificial Immune Algorithm for the project scheduling problem under resource constraints. *Applied Soft Computing*, *11*(2), pp.1975-1982.
[46] :Silaparasetty, N. and Silaparasetty, N., 2020. Introduction to Jupyter Notebook. Machine Learning Concepts with Python and the Jupyter Notebook Environment: Using Tensorflow 2.0, pp.91-118.

Annexe

En 1998, les laboratoires de MIT Lincoln ont organisé un programme d'évaluation des systèmes de détection d'intrusion DARPA dans le but d'examiner et d'évaluer les recherches dans la détection d'intrusion. Pour cela, ils ont installé un environnement pour acquérir des connexions TCP/IP pendant neufs semaines dans un réseau local (LAN) simulant un LAN typique de l'Armée de l'Air des États-Unis. Le réseau a été actionné mais en lui injectant de multiples attaques. Pour chaque connexion TCP/IP, 41 attributs quantitatifs et qualitatifs ont été extraits (voir Annexe A). La compétition "KDD Intrusion Detection 1999" utilisait un sous ensemble de 494021 enregistrements ce qui représente 10% de la base de données globales.

- Probing: surveillance et sondage
- DOS (Denial Of Service): déni de service
- R2L (Remote to User): accès non autorisé à partir d'une machine distante
- U2R (User to Root): accès non autorisé pour avoir le privilège d'un administrateur

1.1 Les attributs KDD99

Les données d'audit du trafic réseau brutes (paquet IP par exemple) ne sont adaptées à la détection d'intrusion. La construction d'attributs était ainsi nécessaire pour extraire l'ensemble d'attributs qui permet de détecter efficacement les intrusions. Il existe trois types d'attributs dans chaque enregistrement de l'ensemble de données KDD99:

Attributs intrinsèques ou de base (Intrinsic features) : ces attributs décrivent les informations de base d'une connexion, telles que la durée, les hôtes source et destination, port et flag.

Attributs du trafic (Traffic features) : ces attributs sont basés sur des statistiques, tels que le nombre de connexions vers la même machine.

Attributs du contenu (Content features) : ces attributs sont construits à partir de la charge utile (Data) des paquets du trafic tels que nombre d'échec de connexion, si connecté en tant que root, et le nombre d'accès au fichier de contrôle.

Chaque enregistrement qui représente une connexion dans l'ensemble de données KDD99 est constitué donc de 41 attributs et une seule valeur cible indiquant le nom d'attaque.

Distribution des attaques dans l'ensemble de données KDD99

La répartition des attaques dans l'ensemble d'apprentissage/test KDD99 est présentée dans les tableaux suivants :

Classes	Normal	Probe	DOS	R2L	U2L	Total
Nombre	97278	4107	391458	1126	52	494021
Pourcentage	19.69%	0.8313%	79.24%	0.2279%	0.0105%	100%

Tableau -Répartition des attaques dans l'ensemble d'apprentissage KDD99

Classes	Normal	Probe	DOS	R2L	U2L	Total
Nombre	60593	4166	229853	16189	228	311029
Pourcentage	19.48%	1.34%	73.90%	5.20%	0.0733%	100%

Tableau -Répartition des attaques dans l'ensemble de test

La distribution des attaques dans l'ensemble d'apprentissage et dans l'ensemble de test n'est pas identique. L'ensemble de test contient plus les 22 types d'attaque d'apprentissage 14 autres types. L'ajout de nouvelles attaques dans l'ensemble de test permet de rendre la tâche de détection beaucoup plus réaliste .

D'après les tableaux, vous constatez que l'ensemble de données KDD99 est un ensemble déséquilibré. En d'autres termes, certaines intrusions produisent plus de connexions que d'autres. Nous sommes donc fasse un problème d'intrusions déséquilibrées.

Avec un tel ensemble de données, l'algorithme d'apprentissage essaie de minimiser le taux d'erreur global en diminuant le taux d'erreur des classes majoritaires (telle que DOS) et en augmentant le taux d'erreur des classes minoritaires (R2L et U2L) ce qui influe sur le taux de détection des intrusions minoritaires. Nous devons alors améliorer le taux de détection des intrusions minoritaires tout en maintenant un taux de détection global raisonnable.

Il existe une solution à ce problème. C'est d'utiliser la technique d'échantillonnage "Sampling techniques" : "over-sampling " c'est-à-dire augmenter le nombre des intrusions minoritaires et "down-sampling" diminuer le nombre des intrusions majoritaires. Puisque, le trafic réseau est énorme, la réduction du nombre des intrusions majoritaires (exemple : trafic normal et Déni de service) peut accélérer considérablement la construction du modèle en réduisant la taille de

l'ensemble de données. L'augmentation du nombre des intrusions minoritaires (exemple : U2R et R2L), quant à elle, peut amplifier leurs poids pour diminuer leurs taux d'erreur.

2. Architecture du système

Le système IDS applique les techniques du datamining pour la construction d'un modèle de détection d'intrusion réseaux. L'architecture du NIDS proposé est donnée dans la figure suivante. On trouve deux phases dans cette architecture : la phase off-line et la phase online.

Dans la première phase le système génère son modèle d'intrusion et dans le deuxième il détecte les intrusions

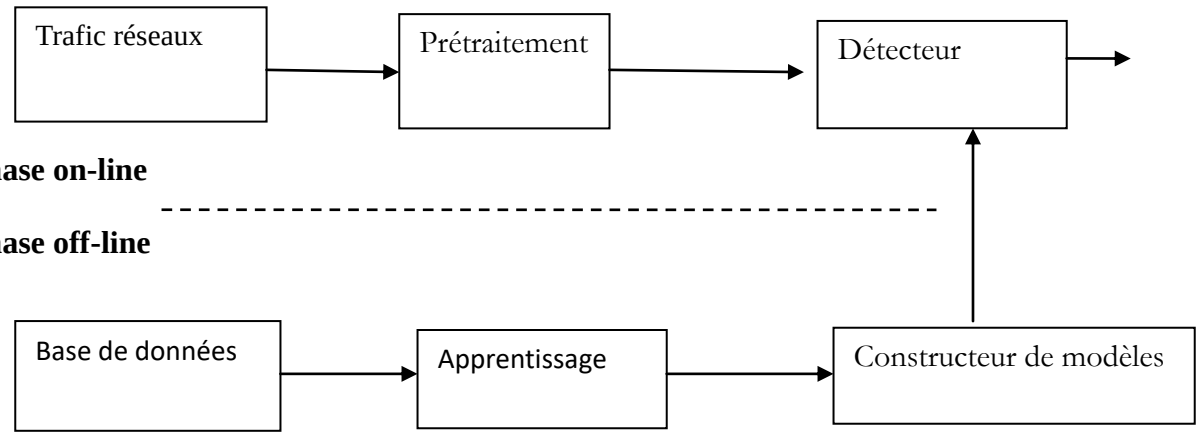


Figure : fonctionnement général du NIDS

Phase on-line

Elle consiste à mettre en place le NIDS développé. Dans cette phase le rôle du système sera alors d'analyser le trafic réseau capté et de détecter en temps réel à l'aide du modèle d'intrusion généré, les intrusions visant le réseau.

Phase off-line

Cette phase nécessite de recueillir suffisamment de données historiques qui incluent à la fois les connexions réseau normales et anormales (attaques).l'ensemble de données KDD99 est en y un bon choix. Il est constitué de données captées à l'aide du sniffer Tcpdump**.**

3 Fonctionnement

Nous disposons de deux ensembles de données KDD99 : l'ensemble d'apprentissage et l'ensemble de test dont nous trouvons deux types de données : données normales et les données représentant des attaques. Dans cette section nous allons présenter les différents modules et la participation des différents ensembles de données à la construction des différents modèles de détection d'intrusion.

L'ensemble de données KDD99 (apprentissage et test) est fournit sous forme de deux fichiers texte.

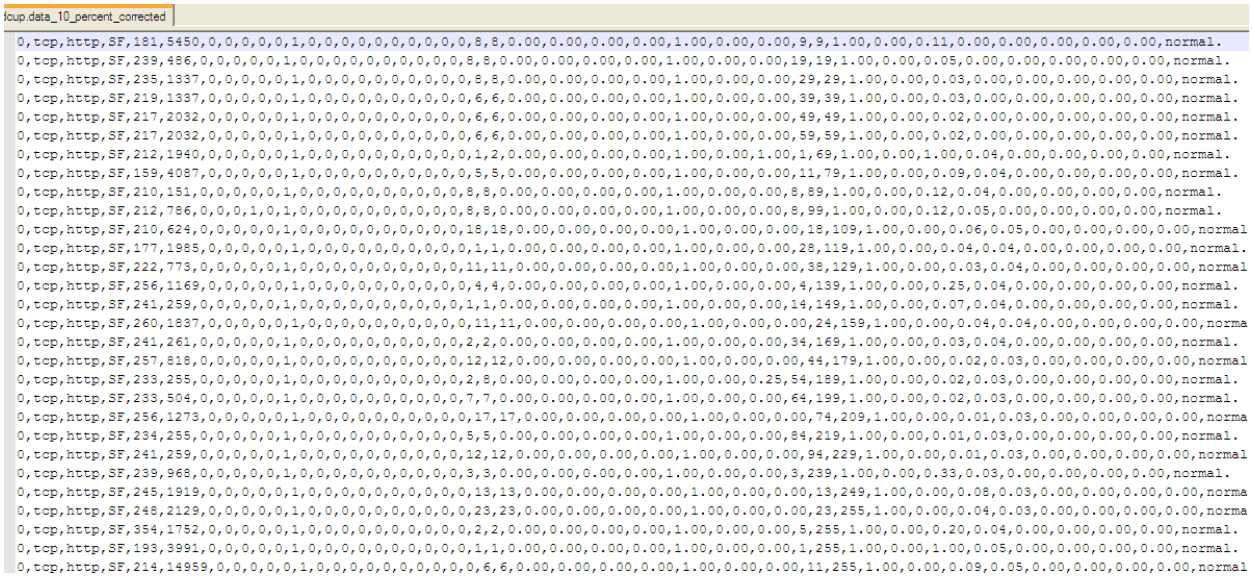


Figure: base de données KDD99 fichier texte

Module de prétraitement de données

Le rôle de ce module est de préparer les données pour qu'elles soient directement exploitables par les différents modules de traitement (apprentissage, validation et classification). $$Les attributs de l'ensemble de données KDD99 sont un mélange d'attributs continus, discrets et symboliques avec d'intervalles de valeurs très variés. La plupart des techniques de classification ne sont pas capable de traiter un tel format de données. Pour cela le prétraitement est nécessaire avant la construction du modèle de classification. Il est constitué de plusieurs étapes qui s'exécutent successivement (ou alternativement si nécessaire) :

Étape d'élimination de redondances :

Après la suppression des enregistrements dupliqués, le nombre total d'enregistrements dans l'ensemble d'apprentissage original et dans l'ensemble final est donnée dans le tableau suivant :

Nombre classe	**Normal**	**Probe**	**Dos**	**R2L**	**U2R**	**Total**
Ensemble originale	97278	4107	391458	1126	52	494021
Ensemble final	87832	2131	54778	999	52	145792

Tableau Nombres d'enregistrements avant et après élimination des enregistrements dupliqué

Numérisation de données

Cette étape consiste à convertir les attributs symboliques en numériques. La procédure est comme suit :

En premier, Les noms d'attaques (comme buffer_overflow, guess_passwd, etc.) sont rangées dans une des cinq classes : 0 pour Normal, 1 pour Probe, 2 pour DOS, 3 pour U2R et 4 pour R2L. Les attributs symboliques comme protocol_type (3 symboles différents), service (70 symboles différents) et flag (11 symboles différents) sont transformés à des valeurs entières de 0 à N-1 où N est le nombre de symboles

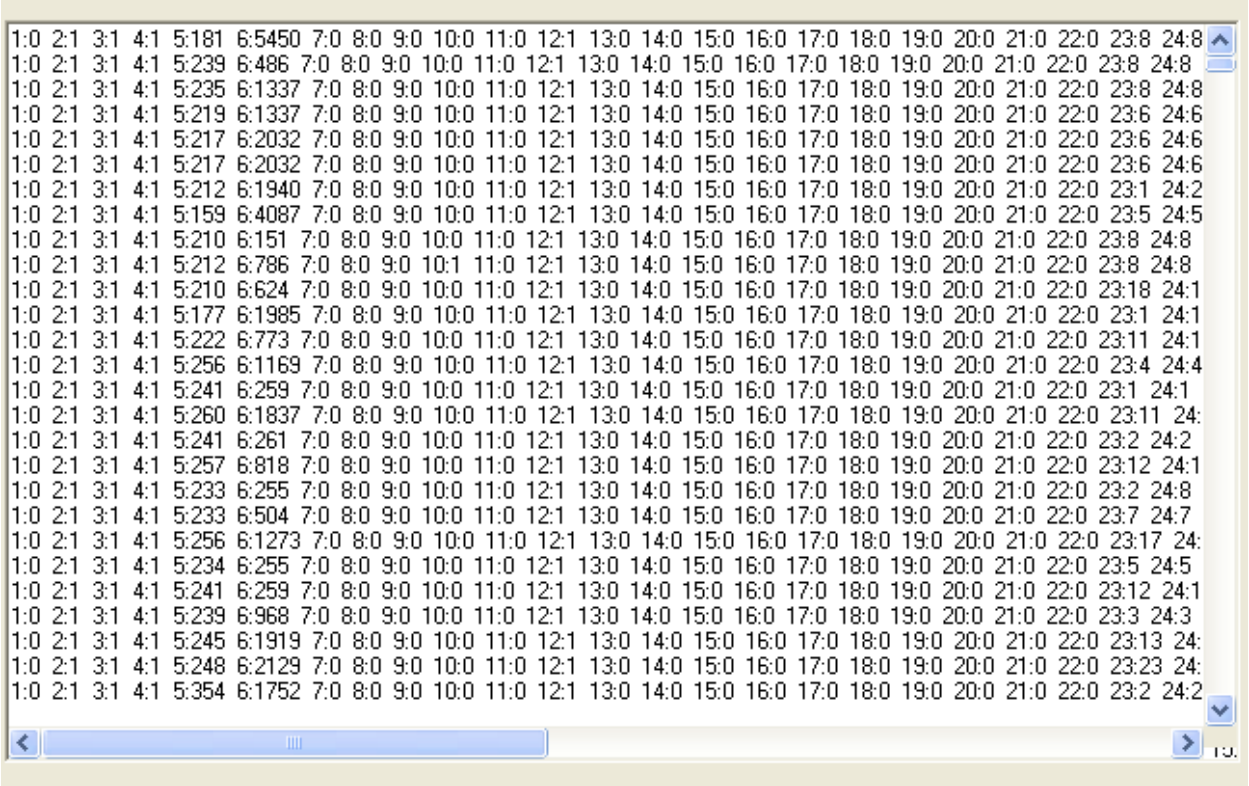


Figure : données après numérisation.