**UNIVERSITE LIBRE DES PAYS DES GRANDS LACS**

**FACULTE DE SCIENCES ET TECHNOLOGIES**

**DEPARTEMENT DE GENIE INFORMATIQUE**



BP. 368 GOMA

[www.ulpgl.net](http://www.ulpgl.net)

**CONCEPTION ET REALISATION D’UN SYSTEME DE SURVEILLANCE D’UN RESEAU INFORMATIQUE**

Par**: BULONZA BADESIRE Lydie**

Travail présenté en vue de l'obtention du Diplôme de Licence en Sciences de l’ingénieur

**Option** : Génie Informatique

**Directeur**: Prof. Ing. Dr Alain AKWIR NKIEDIEL

**Encadreur**: MSc. Ing. Kambale WAMUHINDO Abednego

**Année académique 2022 - 2023**

Epigraphe

« We cannot be certain of being right about the future; but we can be almost certain of being wrong about the future, if we are wrong about the past. »

**Gilbert Keith Chesterton, English writer, poet, philosopher, dramatist, journalist, orator, lay theologian, biographer, and literary and art critic**

Dédicace

Cette partie n’est pas obligatoire mais vous pouvez y faire une dédicace de votre travail à un individu (pas plus) ou à un groupe d’individus.

Remerciements

Ici vous pouvez adresser vos remerciements à différents individus ou différentes organisations/institutions pour des raisons que vous devez spécifier en rapport avec votre travail de recherche. (**NE PAS DEPASSER UNE PAGE**)

Résumé

*Présentation en quelques lignes du contenu du mémoire. Mais ne pas dépasser 1 page. Présenter les principaux objectifs des recherches, les approches utilisés, les principaux résultats obtenus et leur importance.*

*NB :*

* *Si votre résumé n’est pas clair et intéressant, les lecteurs peuvent abandonner la lecture de votre travail ;*
* *La clarté du langage reflète la clarté des idées ;*
* *Tout au long de votre rédaction, il est important de fournir beaucoup d’effort afin de bien écrire car beaucoup de fautes peuvent fatiguer rapidement le lecteur et l’énerver à tel point qu’il peut mal juger tout le travail (même si le fond est bon).*
* *Votre rôle en tant qu’auteur est de communiquer des informations et des connaissances de manière attrayante et bien écrite.*
* *Utiliser toujours un bon français tout le long du manuscrit, et ne pas oublier les légendes pour les figures, les photos, les graphiques, et les tableaux.*

**Abstract**

Ici vous devez mettre la traduction anglaise du sommaire ou résumé. Au cas vous rencontrez des difficultés avec l’anglais, vous devez solliciter de l’aide et non vous contenter d’une traduction par Google translate qui souvent n’est pas très bonne.

Table des matières

[Epigraphe 1](#_Toc166578727)

[Dédicace 2](#_Toc166578728)

[Remerciements 3](#_Toc166578729)

[Résumé 4](#_Toc166578730)

[Table des matières 6](#_Toc166578731)

[Liste des abréviations 8](#_Toc166578732)

[Liste des tableaux 9](#_Toc166578733)

[Liste des figures 10](#_Toc166578734)

[0. Introduction générale 12](#_Toc166578735)

[0.1. Généralités sur le thème 12](#_Toc166578736)

[0.2. Identification et formulation du problème 13](#_Toc166578737)

[0.3. Questions de recherche 13](#_Toc166578738)

[0.4. Formulation des hypothèses 14](#_Toc166578739)

[0.5. Justification du choix du sujet et motivations 14](#_Toc166578740)

[0.6. Énoncé des objectifs de recherche 15](#_Toc166578741)

[0.6.1. L’objectif général 15](#_Toc166578742)

[0.7. Méthodologie et délimitation du travail 16](#_Toc166578743)

[0.8. Structure du mémoire/ Subdivision du travail 16](#_Toc166578744)

[Chapitre 1 Généralités sur le réseau et la sécurité des réseaux informatiques 17](#_Toc166578745)

[1.1 Introduction 17](#_Toc166578746)

[1.2 Généralités des réseaux informatiques 17](#_Toc166578747)

[1.2.1 Définitions et terminologie 17](#_Toc166578748)

[1.2.2 Type de réseaux informatiques 18](#_Toc166578749)

[1.2.3 Communication sur un réseau 21](#_Toc166578750)

[1.2.4 Protocole réseau 23](#_Toc166578751)

[1.3 Trafic et attaques courants 24](#_Toc166578752)

[1.3.1 Trafic et adressage réseau 24](#_Toc166578753)

[1.3.2 Attaques courant sur le réseau 25](#_Toc166578754)

[1.4 Les techniques courantes de détection des attaques réseaux [2] [17] [18] 27](#_Toc166578755)

[Chapitre 2 Analyse et conception 31](#_Toc166578756)

[**2.1** **Introduction** 31](#_Toc166578757)

[**2.2** **Conception de l’architecture globale** 31](#_Toc166578758)

[**2.2.1** **Spécifications du système** 31](#_Toc166578759)

[**2.2.2** **Architecture du module de détection d’anomalies** 32](#_Toc166578760)

[**2.2.3** **Architecture du module de surveillance réseau** 33](#_Toc166578761)

[**2.3** **Présentation des diagrammes** 33](#_Toc166578762)

[**2.3.1** **Diagramme de cas d’utilisation** 33](#_Toc166578763)

[**2.3.2** **Diagramme de séquences** 38](#_Toc166578764)

[**2.4** **Présentation des interfaces** 44](#_Toc166578765)

[**Conclusion partielle** 47](#_Toc166578766)

[Chapitre 3 Implémentation et présentation des résultats 48](#_Toc166578767)

[3.1 Introduction 48](#_Toc166578768)

[3.2 Présentation du dataset 48](#_Toc166578769)

[3.3 Pretraitement des données 48](#_Toc166578770)

[3.3.1 Feature selection 49](#_Toc166578771)

[Bibliographie 55](#_Toc166578772)

Liste des abréviations

**UDP:** User Datagram Protocol

**IP:** Internet Protocol

**TCP**: Transmission Control Protocol

**IEEE**: Institute of Electrical and Electronics Engineers

Liste des tableaux

[Tableau 1 : Tableau présentant les types d’apprentissages et les algorithmes associés 29](#_Toc166578808)

[Tableau 2: tableau de verbes des cas d'utilisation 34](#_Toc166578809)

[Tableau 3: documentation du cas d'utilisation "créer Compte » 35](#_Toc166578810)

[Tableau 4: documentation du cas d'utilisation "se Connecter" 35](#_Toc166578811)

[Tableau 5: documentation du cas d'utilisation "configurer Système" 36](#_Toc166578812)

[Tableau 6: documentation du cas d'utilisation "récolterTrafic" 36](#_Toc166578813)

[Tableau 7: documentation du cas d'utilisation "analyserTrafic" 37](#_Toc166578814)

[Tableau 8: documentation du cas d'utilisation "alerter" 37](#_Toc166578815)

[Tableau 9: documentation du cas d'utilisation "se Déconnecter' 38](#_Toc166578816)

Liste des figures

[Figure 1:point d'accès de la marque D-Link 18](#_Toc166579427)

[Figure 2: Modèle OSI 21](#_Toc166579428)

[Figure 3:Suite des protocoles TCP/IP [13]. 23](#_Toc166579429)

[Figure 4: Diagramme de cas d'utilisation 33](#_Toc166579430)

[Figure 5: Diagramme de séquence pour le cas d'utilisation "créerCompte" 38](#_Toc166579431)

[Figure 6: Diagramme de séquence pour le cas d'utilisation "se Connecter" 39](#_Toc166579432)

[Figure 7: diagramme de séquence pour le cas d'utilisation "configurer Système" 40](#_Toc166579433)

[Figure 8: diagramme de séquence pour le cas d'utilisation "capturer Trafic" 41](#_Toc166579434)

[Figure 9: Diagramme de séquence pour le cas d'utilisation "analyserTrafic" 41](#_Toc166579435)

[Figure 10: diagramme de séquence pour le cas d'utilisation "alerter" 42](#_Toc166579436)

[Figure 11: diagramme de séquence pour le cas d'utilisation "seDeconnecter" 42](#_Toc166579437)

[Figure 12: maquette de l'interface de création de compte utilisateur 44](#_Toc166579438)

[Figure 13: maquette de l'interface de connexion de l’utilisateur 45](#_Toc166579439)

[Figure 14: maquette de l'interface globale du système 46](#_Toc166579440)

1. Introduction générale
   1. Généralités sur le thème

Les réseaux sont à ces jours omniprésents dans la vie quotidienne et dans le monde du travail et sont devenus des éléments centraux de tous les systèmes d’information, des entreprises, des industries et des institutions. Ils simplifient la communication et le partage des données en offrant une interconnexion. Cette facilitation n’est toutefois sans risque et exige la mise en place des mécanismes de sécurisation et de gestion, afin d’offrir aux personnes et systèmes qui sont connectés au réseau, une expérience optimale et leur garantir la sécurité [1] [2].

La sécurité réseau est un processus continu qui inclue la mise en place des structures de sécurité, leur affermissement et leur surveillance, répondre aux attaques et déterminer les attaquants. Les professionnels de la sécurité sont souvent appelés à suivre tout ce processus à la loupe et doivent comprendre les menaces auxquels leur système est confronté. La mise en place de ces mécanismes peut s’avérer être une tache fastidieuse car elle exige la surveillance de différents composants et une minutie dans l’exécution des taches [1] .Afin de bien mener à bien ces tâches, plusieurs outils entrent en jeu afin de simplifier le processus tout en garantissant une protection plus ou moins maximale, notamment les pare feux, les systèmes de détections d’intrusion sur le réseau (N-IDS), et tant d’autres. Il est primordial de disposer d’outils de sécurité pour le réseau informatique pour garantir la sécurité des personnes et des leurs données qui transitent par le réseau informatique.

* 1. Identification et formulation du problème

Avec l’intégration croissante des réseaux dans le quotidien, la sécurité de ce dernier est devenue de plus en plus complexe à assurer.

* Les données sensibles se retrouvent exposées à des actes de malveillance dont la nature et les méthodes sont sans cesse changeantes,
* Les risques se multiplient et se modifient, et l’ingéniosité des attaqueurs est grandissantes, chaque mois il y a des nouvelles formes d’attaques (les zeros-days attaques) qui la plupart de temps ne sont identifiées qu’après la réussite de l’attaque.
* La détection des anomalies sur le réseau est une tache accablante car les systèmes utilisés génèrent beaucoup d’alertes dont la majeure partie est constituée des faux positifs et des faux négatifs, ce qui ne permet pas à celui qui surveille le réseau d’orienter les efforts au bon endroit.

La surveillance d’un réseau s’avère être une tâche qui exige des outils efficaces afin de faciliter la tâche aux personnels de sécurité réseau et de leur assurer une efficacité maximale.

* 1. Questions de recherche

Dans le but d’apporter une solution aux problèmes soulevés précédemment, nous essayeront de répondre aux questions ci-après :

* Comment développer une solution adaptative qui soit en mesure de détecter toutes les anomalies sur le réseau afin d’accroitre la sécurité ?
* Comment minimiser les alertes faux-négatifs et faux-positifs renvoyé par un système de détection d’anomalies sur un réseau ?
* Quels sont les éléments nécessaires à tenir en compte pour administrer efficacement un réseau, garantir la sécurité des données et réduire les risques d’attaques ?
  1. Formulation des hypothèses

Pour mener à bout cette recherche et apporter des réponses aux questions formulées, nous supposons que :

L’utilisation de l’intelligence artificielle, notamment le machine Learning (avec la technique d’apprentissage supervisé) pour la détection des anomalies pourrait permettre de détecter un nombre plus élevé d’attaques comparé aux méthodes primaires, et ainsi rendre le système plus performant et évolutif

Le choix d’un algorithme de machine Learning bien ajusté permettrait également d’accroitre la précision du système et permettrait ainsi de générer moins d’alertes faux-négatifs et faux-positifs. Nous supposons également qu’offrir au personnel de sécurité réseau une interface de visualisation présentant les éléments les plus pertinents du réseau tel que l’utilisation de la bande passante, la surveillance des services utilisés et du trafic lui permettrait de mieux administrer le réseau et pouvoir intervenir plus facilement en cas de problème.

* 1. Justification du choix du sujet et motivations

La motivation primaire de ce travail, vient du fait de la nécessité d’acquérir une connaissance spécifique dans les domaines qui font objets d’un énorme intérêt scientifique pour moi, notamment le domaine des réseaux et de la sécurité informatique.

Aussi la motivation vient du fait que le problème de la sécurité croît avec le développement et toutes les avancées technologiques que nous avons actuellement notamment l’internet des objets et l’intelligence artificielle. Les données sont devenues la matière abondante de toutes les expériences et plusieurs moyens sont mis en œuvre pour les récolter, il est impérieux de penser à des solutions de sécurité pour les données des personnes tant physique que morale.

Au terme de cette recherche nous espérons proposer une solution convenant aux enjeux socioéconomiques de notre société ou du moins aboutir à une piste de solution qui pourrait être utilisé pour des futures études.

* 1. Énoncé des objectifs de recherche
     1. L’objectif général

L’objectif générale de cette recherche est de parvenir à développer une solution de surveillance du trafic sur un réseau informatique ; développer un outil de détection d’intrusion offrant précision et rapidité tout en minimisant les alertes faux-positifs et les faux-négatifs afin d’aider les administrateurs de réseau.

* + 1. **Les objectifs spécifiques**

Des objectifs spécifiques, nous pouvons citer entre autres :

* Offrir une compréhension globale de la sécurité informatique.
* Etudier les méthodes couramment utilisées dans la surveillance des réseaux informatiques
* Décrire le processus de la manière dont le trafic se comporte sur un réseau informatique
* Etudier l’efficacité d’un algorithme de machine Learning dans la détection des intrusions
  1. Méthodologie et délimitation du travail

La méthodologie de travail utilisé dans le déroulement de cette recherche est la déduction et l’expérimentation. Nous avons effectué la récolte des données, leur analyse et enfin la conception du système et le test.

Les données utilisées dans ce travail proviendront de bases des données existantes telles que le dataset NSL-KDD et le KDD-99 qui contiennent des données d’entrainement et de test.

Le système développé est adapté pour les réseaux sans-fils, et ne permettra de détecter qu’un trafic est soit normal ou anormal. Ce système constitue un outil qui ne sera pas à lui-même suffisant pour administrer un réseau mais il pourra aider l’administrateur du réseau à effectuer des tâches telles que la gestion des utilisateurs, la surveillance du trafic ainsi que l’affichage des alertes provenant du détecteur d’intrusion (NIDS).

* 1. Structure du mémoire/ Subdivision du travail

A part l’introduction et la conclusion, notre travail est ainsi constitué :

1. Du chapitre premier intitulé *Généralités sur le réseau et la sécurité des réseaux informatiques.* Nous donnons une explication succincte des réseaux informatiques et leur sécurité.
2. Du chapitre deuxième intitulé *Analyse et conception du système***,** ici nous présentons la conception et la réalisation de notre système.
3. Du chapitre troisième intitulé *Réalisation et discussions des résultats* dans lequel nous décrivons l’implémentation du système et présentons la phase de test du système et la présentation des résultats.

# Généralités sur le réseau et la sécurité des réseaux informatiques

## Introduction

Dans ce chapitre nous allons aborder les généralités des réseaux informatiques. Nous donnerons dans ce chapitre les informations nécessaires pour la compréhension des réseaux et nous présenterons également les attaques courant auxquels un réseau peut être soumis tout en montrant les techniques utilisés à ce jour afin de détecter ces attaques.

## Généralités des réseaux informatiques

### Définitions et terminologie

Guy Pujolle dans son livre *Initiation aux réseaux : cours et exercices,* définit un réseau comme étant un ensemble d’équipements et de liaisons de télécommunications autorisant le transport d’une information, quelle qu’elle soit, d’un point à un autre, où qu’il soit.

Le but du réseau est de permettre l’interconnexion de différents équipements afin de faciliter le partage, il permet le partage des ressources d’un hôte vers un autre [3].

La topologie du réseau représente l’agencement des équipements entre eux : des réseaux peuvent être organisés en boucle, en arborescence, en mailles, etc. Afin que ces derniers puissent communiquer entre eux, il est nécessaire d’une part qu’elles sachent exploiter un canal de communication adapté (des câbles électriques ou optiques, des ondes radio, la lumière infrarouge...), mais aussi et surtout qu’elles soient capables de se synchroniser et de se comprendre. Pour cela, des règles de communication ont été définies. Ces règles sont appelées protocole de communications nous en parlerons dans un point y consacré [5].

### Type de réseaux informatiques

Les réseaux peuvent être classés selon différentes critères, notamment :

* Selon la transmission des données nous avons les réseaux filaires et les réseaux sans fils
* Selon la taille et l’échelle géographique ; nous distinguons : Personal Area Network (PAN) ou réseau personnel, Local Area Network (LAN) ou réseau local, Metropolitan Area Network (MAN) ou réseau métropolitain, Wide Area Network (WAN) ou réseau étendu et le Global Area Network (GAN) ou réseau global

Dans ce travail nous nous focaliserons sur les réseaux sans fil.

#### Réseau sans fil

Un réseau sans fil ou Wireless Network est un réseau informatique numérique qui connecte différents postes ou systèmes entre eux par ondes radio. Ces réseaux ont été conçus afin de répondre aux inconvénients du câblage car les systèmes de câblages sont complexes et couteux.

Les réseaux sans fil sont normalisés et fonctionnent suivant différents protocoles qui définissent d’une part la gestion de la bande passante, le format des trames et d’autres parts les procédures d’échanges.

Ils peuvent être classés en trois catégories à savoir :

* WPAN pour les micro-réseaux personnels
* WLAN pour les réseaux Ethernet locaux
* WMAN pour les réseaux à grande échelle.

#### Le Wi-Fi

Wi-Fi ou « Wireless fidélité » est le nom commercial des standards WLAN du groupe 802.11 défini par l’IEEE. Le 802.11 est une norme définissant le protocole de communication sans fil entre différents périphériques informatiques. [6] [7] [8]

#### Le point d’accès sans fil

Un point d’accès sans fil (AP) permet aux dispositifs sans fil de se connecter à un réseau sans fil.  Il prend la bande passante provenant du routeur et l’étire pour que plusieurs dispositifs puissent se connecter au réseau à des distances plus importantes. Un point d’accès sans fil ne se contente pas d’étendre seulement la portée de l’accès Wi-Fi mais aussi, il fournit également des informations utiles à propos des dispositifs sur le réseau.

Généralement il s’agit d’un routeur sans fil, il allie les fonctions de réseau d’un point d'accès sans fil et d’un routeur. [8]



Figure 1‑1:point d'accès de la marque D-Link

#### SSID

SSID est l'abréviation de l'anglais "service set identifier" (ou identifiant défini de service), qui est un identifiant important en ce qui concerne les réseaux sans fil. Concrètement, un SSID est le nom attribué à un réseau Wi-Fi lors de la configuration du routeur. En somme, un SSID est un simple nom attribué à un réseau et qui sert à l'identifier.

Le SSID agit ensuite comme le point d'accès afin que les appareils proches, comme les smartphones ou les ordinateurs portables, puissent le localiser et s'y connecter. Lorsqu'ils essaient de connecter leur appareil, les utilisateurs peuvent voir une liste de réseaux Wi-Fi disponibles et utiliser le SSID comme un nom afin de distinguer le réseau qu'ils souhaitent utiliser. L'identifiant est souvent enregistré par l'appareil, de sorte que ce dernier peut par la suite se reconnecter automatiquement au réseau. [9]

#### La bande passante et débit

Le terme bande passante fait généralement référence au volume d’informations par unité de temps qu’un support de transmission peut gérer. Il mesure la quantité de données pouvant être envoyées sur une connexion spécifique dans un laps de temps donné. C’est la quantité de donnée par unité de temps.

Le débit en revanche est utilisé pour mesurer la vitesse de téléchargement soit donc, c’est le taux de transfert. [10]

#### Interface réseau

Une interface réseau désigne généralement le point de connexion physique ou logique entre un dispositif (comme un ordinateur, un routeur, un commutateur, etc.) et un réseau informatique. En d’autres termes, une interface réseau est la partie qui assure la connexion entre un terminal utilisateur et un réseau. Cette interface peut être de différents types, selon le moyen de connexion au réseau, comme :

1. Carte réseau (NIC - Network Interface Card) : elle permet la connexion physique à un réseau à l'aide de câbles Ethernet, Wi-Fi, ...
2. Interface sans fil (Wi-Fi) : qui permet de se connecter au réseau sans fil (Wi-Fi).
3. Interface Bluetooth : qui permet la communication sans fil à courte portée.
4. Interface Ethernet : Permet la connexion filaire.
5. Interface virtuelle : Créée par des logiciels pour simuler plusieurs interfaces physiques sur une même carte réseau matérielle.
6. Interface de boucle locale (Loopback) : Utilisée pour les tests et le diagnostic réseau en permettant à un ordinateur de s'envoyer des données à lui-même.

### Communication sur un réseau

Lorsque deux appareils sont interconnecté et communique, les informations qu’ils se partagent passent par différents niveau avant d’atteindre la destination. En parallèle différentes transformations sont appliquées sur la dite information afin de faciliter son acheminement. Un modèle international a été développé par l’ISO qui définit la structure de ces niveaux que nous appellerons couches, ainsi que la fonction des protocoles de communication de données associées à chaque couche. Ce modèle est appelé modèle OSI. Il est constitué de sept couches [9], qui sont :

1. La couche physique : offre tous les matériels et logiciels nécessaires au transport correct des informations.
2. La couche de liaison : permet la reconnaissance du début et de la fin de l’information à envoyer (trame) afin que celle-ci soit transmise et capté correctement.
3. La couche réseau : a pour but d’acheminer correctement les paquets d’information jusqu’au récepteur.
4. La couche de transport : assure le transport des messages de l’émetteur jusqu’au récepteur en fragmentant le message en paquets puis à réassembler ces paquets à la réception pour retrouver le message de départ.
5. La couche de session : permet l’ouverture et la fermeture des sessions entre les utilisateurs.
6. La couche de présentation : cette couche se charge de la présentation de l’information, elle met en forme les données pour les rendre compréhensibles par le destinataire.
7. La couche d’application : fournit la possibilité de s’échanger des informations par le réseau sous-jacent.

La figure 3 présente l’organisation du modèle OSI. Les couches basses s’intéressent au transport de l’information, tandis que les couches hautes correspondent à leur traitement. Chaque couche a des protocoles associés qui réalisent les fonctionnalités associées à cette dernière.

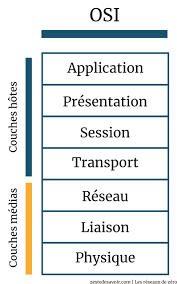


Figure 1‑2: Modèle OSI

Le modèle OSI a été conçu pour etre le modèle standard pour la communication des données mais cela n’a pas été le cas. La suite de protocole TCP/IP est devenue l’architecture commerciale la plus dominante. A ce jour, le modèle TCP/IP constitue la base de la communication Internet [13].

Différemment au modèle OSI le modèle TCP/IP est constitué de cinq couches dont :

La couche physique, la couche de liaison, la couche réseau, la couche de transport et la couche application [10].

### Protocole réseau

Plus haut dans ce chapitre nous avons mentionné le protocole. En fait, pour que la communication entre les appareils connectés sur un réseau soit possible, ces derniers doivent utiliser les mêmes règles de communication, de syntaxe, de synchronisation et de sémantique que l’on appelle protocoles. En des termes simples, on peut dire que le protocole c’est le langage qui permet aux équipements interconnecté de se comprendre et communiquer.

Chaque couche de l’architecture de communication utilisée a des protocoles prédéfinies qui permettent aux données de transiter au travers de la couche.

La figure 4 présente les protocoles suivant leur répartition par couche.

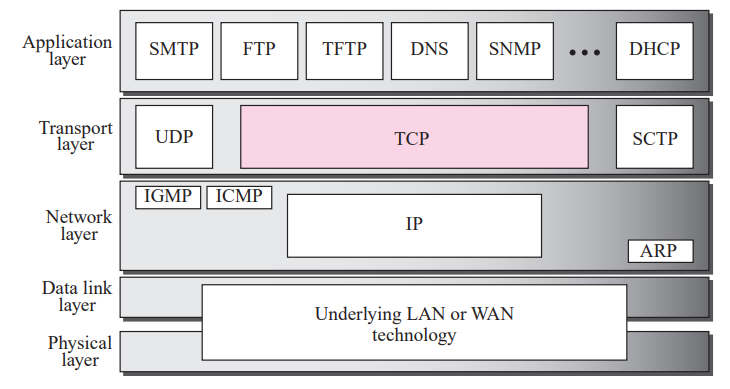


Figure 1‑3:Suite des protocoles TCP/IP [13].

## Trafic et attaques courants

### Trafic et adressage réseau

Considérons deux ordinateurs A et B. ces deux sont connecté sur un même réseau et souhaitent échanger des données. Lorsque la machine A envoie par exemple une image à la machine B, en fait cette image est d’abord fragmenté en plusieurs fragments, puis envoyé séparément. Chaque fragment comprend une entête qui permet de savoir ou l’envoyer et dans quel ordre faire la reconstitution du paquet à l’arrivé. Suivant la couche sur laquelle on se trouve ces fragment porte de nom différent respectivement datagramme, segment, paquet, trame, bits pour la couche application, transport, réseau, liaison et la couche physique [13].

Ce flux de fragment quittant d’un équipement vers un autre constitue ce que l’on appelle trafic.

Le trafic est rendu possible par le fait que chaque appareil connecté sur le réseau a des identifiants, ceux-ci permet d’identifier les appareils. Il s’agit de l’adresse IP et l’adresse MAC.

L’adresse MAC est un identifiant matériel unique attribué à la carte réseau lors de la fabrication de l’appareil. C’est un peu comme son empreinte digitale.

L’adresse IP quant à lui est un nom symbolique unique qui permet d’identifier l’appareil sur le réseau. Il comprend deux parties : l’identificateur du réseau, où se trouvent l’équipement, et l’identificateur de la machine elle-même (qui a une signification locale à ce réseau) [4] [7].

### Attaques courant sur le réseau

Une attaque réseau est défini comme étant une tentative d’acquérir un accès non autorisé à un réseau ou un système dans le but de voler ou détruire des données, de réaliser des activités malveillantes, ou perturber les opérations de l’organisation [13].

Avec la croissance de l’utilisation des réseaux informatiques et le perfectionnement de ce dernier dans ses performances, les hackers ont développés plusieurs types d’attaques motivées par des intérêts allant de la satisfaction personnelle aux intérêts politique ou économique.

On classifie les attaques en deux catégories suivant que celles-ci ont des conséquences sur la nature des données : l’attaque passive et l’attaque active.

Une attaque est dite *passive* si le hacker acquiert l’accès non autorisé au réseau et parvient à surveiller les activités ou vole des données sensibles dans pour autant effectuer des modifications sur ces données.

On parle alors d’attaque *active* lorsque le hacker accède au réseau ainsi que les données et effectue des modifications sur celles-ci, soit en les supprimant, encryptant ou en les détériorant.

Les attaques les plus courants à ce jour sont présentées ci-dessus de manière non exhaustive:

1. Le virus informatique : C’est l’attaque réseau le plus courant et peut se propager d’un ordinateur à un autre.
2. Le logiciel malveillant : c’est la forme d’attaque la plus dangereuse et peut provoquer plus de dommage. Les attaques essaient d’acquérir des accès au système afin de troubler ou corrompre les données avec des codes malveillant.
3. Vers informatique : c’est un programme malveillant qui se propage d’un ordinateur à un autre en se déplaçant.
4. Le phishing : c’est une forme d’attaque d’ingénierie sociale qui consiste à tromper les utilisateurs légitime en les trompant par des liens ou mail qui semble inoffensif.
5. Le déni de service DoS ou déni distribué de service DDoS
6. Man in the middle
7. Ransomware : ou attaque par rançonnement
8. 5G based attacks
9. SQL Injection

En réalité les attaques réseaux sont groupées en quatre catégories globales qui renferment les attaques citées ci-haut, ce sont :

* **Les attaques de déni de service (DoS):** consiste à rendre les ressources inaccessibles et ainsi faire en sorte que les requêtes légitimes des utilisateurs ne soient pas reçues.
* **Les attaques par sonde :** lors d’une attaque de ce type le hacker scan le réseau afin de récolter les informations par rapport aux faiblesses et vulnérabilités qu’il peut exploiter plus tard pour compromettre le système cible.
* **Les attaques U2R :** pour ce type d’attaque un utilisateur normal tente d’abuser du système pour acquérir des privilèges d’administrateur.
* **Les attaques R2L :** pour ce genre d’attaques un utilisateur distant tente d’accéder une machine locale en tant qu’utilisateur local en envoyant des paquets à la machine cible à partir de internet.

## Les techniques courantes de détection des attaques réseaux [2] [17] [18]

La détection des attaques réseau ou détection d’intrusion est définie comme le fait d’identifier les accès non autorisé, l’abus et la mauvaise utilisation des ressources du réseau par les intrus à l’intérieur ou à l’extérieur du réseau.

La détection d’intrusion permet de détecter des attaques hostiles sur le réseau, de surveiller les activités entrants et sortants, les analyser afin de recueillir des signes d’intrusion.

En matière de détection d’attaques trois méthodes sont couramment utilisé pour leur implémentation :

* La détection d’anomalies
* La détection d’attaque basée sur les signatures d’attaques
* La détection hybride

**La détection d’anomalies :** est une méthode qui consiste à détecter le trafic anormal sur le réseau. Ici, on mesure la déviation du trafic par rapport au trafic qui est défini comme étant normal.

**La détection par signature d’attaques :** cette technique recherche les ressemblances et les signatures des attaques connues dans le trafic du réseau. On utilise donc une base de données contenant les signatures des attaques connues à ce jour puis on le compare avec le trafic afin de déceler les points communs et ainsi détecter ces attaques dans le trafic réel du réseau.

**La détection hybride :** cette méthode combine les deux précédents, en combinant les signatures ainsi que la détection des anomalies dans le trafic du réseau.

La détection d’attaques réseaux est basée sur différents techniques dont les plus courants sont :

1. **Les techniques statistiques** : ces techniques sont utilisées lorsqu’on a une idée sur la relation telle que celle que l’on recherche dans les données et avec laquelle on peut associer les mathématiques afin d’aboutir à des résultats. Les techniques statistiques souvent utilisées sont linéaire, non linéaire, les modèles de Markov ou les estimateurs de Bayes.
2. **Le machine Learning**: est une science et un art qui consiste à programmer la machine à apprendre à partir des données. Utiliser le machine Learning pour détecter les attaques consiste à étudier les algorithmes qui permettent d’entrainer un modèle sur un trafic réseau et apprendre à ce dernier comment détecter un trafic qui est en réalité une attaque.

Trois méthodes d’apprentissage sont possibles avec le machine Learning suivant que les données sont étiquetées ou non, il s’agit de : l’apprentissage supervisé, l’apprentissage semi-supervisé et l’apprentissage non supervisé.

Pour chaque type d’apprentissage, différentes algorithmes sont disponibles, nous pouvons en présenter quelques un sans pour autant être exhaustif :

|  |  |
| --- | --- |
| **Type d’apprentissage** | **Algorithmes** |
| ***Apprentissage supervisé*** | ***K-Nearest Neighbors, Logistic regressio, Support Vector Machines (SVMs), Decision Tree, Random Forest, etc.*** |
| ***Apprentissage non supervisé*** | ***K-Means algorithms, DBSCAN, Hierarchical Cluster Analysis (HCA), Isolation Forest, Principal component Analysis (PCA), Apriori, etc.*** |
| ***Apprentissage semi supervisé*** | ***Utilise des combinaisons d’algorithme d’apprentissage supervisé et semi-supervisé.*** |

Tableau 1 : Tableau présentant les types d’apprentissages et les algorithmes associés

1. **Le data mining** : les techniques de data mining offre la facilité dans la détection des attaques en permettant de résumer les données et de les visualiser plus facilement. L’utilisation du data mining permet de réduire les alertes déclenché par le trafic normal afin que l’administrateur du réseau se focalise sur les vraies attaques, identifier les éléments déclencheur de fausses alarmes, trouver les activités anormales qui cachent des attaques et établir le schéma entre les attaques.

Dans ce travail nous développons un système de détection d’anomalie sur un réseau basé sur le machine Learning en utilisant l’apprentissage supervisé.

**Conclusion partielle**

Dans ce chapitre nous avons présenté la généralité sur les réseaux informatiques, nous avons offert les informations basiques pour comprendre les réseaux et leur fonctionnement ainsi que la manière dont la communication se passe sur le réseau. Nous avons présenté les attaques courant sur les réseaux informatiques ainsi que les techniques courant utilisées pour détecter ces dernières.

Dans le chapitre suivant, nous allons modéliser notre système en suivant le principe de conception du génie logiciel notamment l’analyse des besoins, la présentation ses spécification et la conception.

# Analyse et conception

* 1. **Introduction**

Dans ce chapitre, nous abordons la conception de notre système de surveillance du réseau informatique. Nous expliquerons clairement l’architecture de notre système, détaillant les fonctionnalités du système. Nous présenterons les diagrammes en utilisant le langage de modélisation UML ainsi que les maquettes des interfaces du système.

* 1. **Conception de l’architecture globale**

Notre système est composé de deux modules ; un module pour la surveillance du réseau et un module pour la détection des anomalies basé sur le Machine Learning.

Nous concevons ce système afin de donner à l’administrateur du réseau un outil lui permettant de mieux surveiller son réseau et détecter les anomalies.

Globalement notre système fonctionnera comme illustré sur la Figure 4 suivante :

Choisir l’action

Voir l’utilisation du réseau

Analyser le trafic

Collecter le trafic

Passer ce trafic au modèle

Utilisateur du système

Se connecter au système

Configurer le système

Figure 4 : schéma de fonctionnement global du système

La figure 4 présente le fonctionnement global de notre système. L’utilisateur ouvre le système et crée un compte, une fois le compte créé il se connecte. Après la connexion, il configure le système notamment en entrant les informations sur son réseau tel que l’adresse IP du réseau et le nom des interfaces réseaux à surveiller. Par la suite une interface lui est présenté à partir duquel il peut réaliser quelques actions notamment capturer le trafic réseau, l’analyser et recevoir des alertes en cas de trafic anormal ou afficher les utilisateurs du réseau.

Les points suivants de ce chapitre donnent plus des détails sur le système que nous voulons concevoir.

* + 1. **Spécifications du système**

L’administration d’un réseau informatique est une tâche fastidieuse et nécessite plusieurs ressources, il est difficile de les réunir en seul système. Ainsi le système que nous concevons devra répondre à quelques exigences fonctionnelles. Notamment le système doit :

* Permettre à l’administrateur du réseau de s’authentifier
* Répertorier les utilisateurs du réseau
* Offrir la possibilité à l’administrateur du réseau de déconnecter un utilisateur
* Permettre à l’administrateur de configurer le système
* Permettre à l’administrateur d’analyser le trafic réseau
* Identifier le trafic suspect
* Générer des alertes en cas en cas de trafic anormal
* Générer un rapport de sécurité

Afin de garantir une flexibilité dans l’utilisation et une bonne administration du réseau, notre système devra répondre aux exigences non-fonctionnelles suivantes :

* Fiabilité : l'application doit être robuste et minimiser les risques d’erreurs.
* Gestion d’alertes : l'application doit fournir des messages d'erreur de manipulation clairs et concis.
* Ergonomie : l'application doit être facile à utiliser et adaptée aux besoins des utilisateurs.
* Sécurité : l'application doit être sécurisée afin de ne pas violer la vie privée des utilisateurs du réseau, pour cela l’utilisateur doit s’authentifier avant d’utiliser le système.
* Performance : le système doit fonctionner de manière optimale en fournissant avec le minimum de bug, et offrir un temps de réponse inférieur à 5 secondes.
  + 1. **Architecture du module de détection d’anomalies**

La détection d’anomalie est une technique qui permet en sécurité réseau de surmonter les limitations des systèmes de détections d’intrusions base sur les signatures d’attaques. Dans ce travail, nous allons concevoir un module de détection d’anomalies sur un réseau informatique et pour ce faire nous avons comparé différents algorithmes de classification afin de sélectionner celui qui offre les meilleurs performances, notamment la précision, le temps de prédiction pour la question de détection d’anomalie sur un réseau Informatique.

* + 1. **Architecture du module de surveillance réseau**

La surveillance du réseau informatique est une activité importante pour une bonne administration de ce dernier. En effet celle-ci permet à l’administrateur du réseau d’avoir une vision claire du réseau et d’etre proactif face à certaines attaques qui peuvent se produire sur le réseau. Pour ce faire il doit disposer des outils qui lui permettent de faciliter la tâche et le rendre efficace.

Administrer un réseau est un processus vaste et nécessite plusieurs outils qu’on ne saurait intégrer dans un seul, raison pour laquelle la partie du module de surveillance réseau répondra qu’aux fonctionnalités présentées dans la section portant sur les spécifications.

Nous allons présenter les diagrammes de conceptions pour ce module.

* 1. **Présentation des diagrammes**
     1. **Diagramme de cas d’utilisation**

Le diagramme de cas d’utilisation définit les interactions entre un acteur et le système [5]. L’acteur est représenté par un bonhomme en bâtonnet tandis que le système est représenté par un bloc, le cas d’utilisation quant à lui est représenté par une ellipse contenue à l’intérieur du bloc [24]. Chaque acteur est relié au cas d’utilisation auquel il participe.

* + - 1. **Identification des acteurs**

Un acteur est tout celui qui initie un cas d’utilisation, c’est-à-dire celui qui interagit directement avec le système.

Notre système n’a qu’un seul acteur, qui est l’administrateur du réseau car il est le seul à interagir avec le système.

* + - 1. **Identifications de cas d’utilisations**

Un cas d’utilisation « use case » est l’ensemble de séquences, d’actions qui seront exécuté sur le système. Le lien entre ces séquences d’actions constitue l’objectif métier de l’acteur. Un cas d’utilisation n’est donc pas juste une séquence mais un ensemble de ces derniers. Les cas d’utilisations sont rattachés à l’acteur qui les exécute, cela implique que dans notre système, comme nous avons un seul acteur tous les liens sont directement rattachés à lui. Les cas d’utilisations sont introduits par de verbes d’actions tel le cas dans nos spécifications :

|  |  |
| --- | --- |
| Sujet | Verbe |
| Administrateur | Créer un compte |
| Adminstrateur | Se connecter |
| Administrateur | Configurer le système |
| Administrateur | Récolter le trafic |
| Adminstrateur | Analyser le trafic |
| Administrateur | Alerter |
| Administrateur | Se déconnecter |

Tableau 2: tableau de verbes des cas d'utilisation

* + - 1. **Présentation du diagramme de cas d’utilisation**

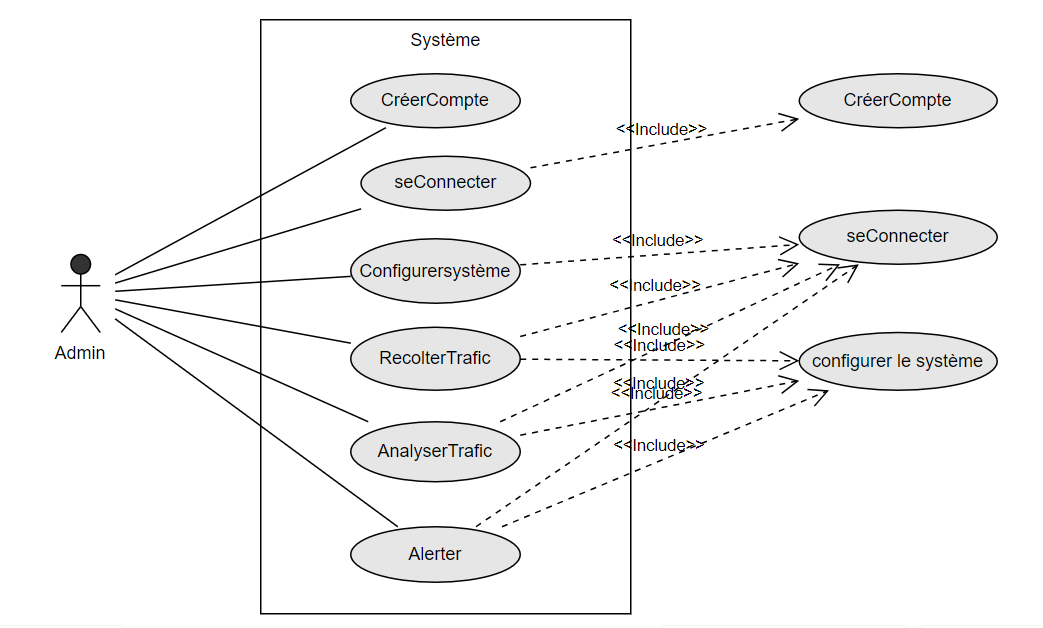
****

Figure 2‑2: Diagramme de cas d'utilisation

* + - 1. **Documentation des cas d’utilisation**
* **Créer un compte**

Tableau 3: documentation du cas d'utilisation "créer Compte »

|  |  |
| --- | --- |
| Cas d’utilisation | Créer un compte |
| Acteurs | Administrateur |
| Préconditions | L’application est ouverte |
| Postcondition | Le Dashboard est visible par l’administrateur |
| Scenario de déroulement | |
| 1. L’administrateur ouvre le système 2. Il charge le formulaire de création de compte 3. Il remplit le formulaire 4. Il clique sur le bouton créer compte 5. Les informations sont enregistrées dans la base de données et le compte est créé. | |

* **Se connecter**

Tableau 4: documentation du cas d'utilisation "se Connecter"

|  |  |
| --- | --- |
| Cas d’utilisation | Se connecter |
| Acteur | Administrateur |
| Précondition | Avoir déjà un compte d’utilisateur |
| Postcondition | Le Dashboard est visible par l’administrateur |
| Scenario de déroulement | |
| 1. L’administrateur charge le formulaire de connexion 2. Il remplit les champs requis 3. Il valide puis le système vérifie si les identifiants sont corrects 4. Si les identifiants sont corrects, il accède au système 5. Sinon il reste sur la page de connexion | |

* **Configurer le système**

Tableau 5: documentation du cas d'utilisation "configurer Système"

|  |  |
| --- | --- |
| Cas d’utilisation | Configurer le système |
| Acteur | Administrateur |
| Précondition | Etre connecté au système |
| Postcondition | Pouvoir utiliser le système |
| Scenario de déroulement | |
| 1. L’administrateur entre les paramètres de configurations notamment : l’adresse IP du réseau et l’interface réseau 2. Il clique sur le bouton « valider » 3. Il peut maintenant utiliser le système | |

* **Capturer le trafic**

Tableau 6: documentation du cas d'utilisation "récolterTrafic"

|  |  |
| --- | --- |
| Cas d’utilisation | Capturer le trafic |
| Acteur | Administrateur |
| Précondition | 1. L’administrateur est connecté au système |
| Postcondition | Le trafic est capturé et enregistré |
| Scenario de déroulement | |
| 1. L’administrateur ouvre le menu trafic 2. Il charge la page de capture de trafic 3. Il clique sur le bouton « capture trafics » 4. Le trafic est capturé et enregistré dans un fichier stocké sur la machine 5. Le trafic non détaillé est affiché sur la page | |

* **Analyser le trafic**

Tableau 7: documentation du cas d'utilisation "analyserTrafic"

|  |  |
| --- | --- |
| Cas d’utilisation | Analyser le trafic |
| Acteur | Administrateur |
| Précondition | L’utilisateur est connecté |
| Postcondition | Le trafic est passé au modèle qui l’analyse et affiche les résultats |
| Scenario de déroulement | |
| 1. L’utilisateur ouvre la page de scan de trafic sous le menu trafic > analyse 2. Il charge le fichier contenant le trafic 3. Le système vérifie si le fichier est du bon format et contient des données valides 4. Si oui les données sont passées dans le modèle de machine Learning qui prédît si le trafic est normal ou anormal 5. Sinon un message d’erreur est renvoyé à l’utilisateur | |

* **Alerter l’utilisateur**

Tableau 8: documentation du cas d'utilisation "alerter"

|  |  |
| --- | --- |
| Cas d’utilisation | Alerter l’utilisateur |
| Acteur secondaire | Administrateur |
| Précondition | Le système est ouvert  Un compte d’utilisateur est connecté  Le trafic est capturé et analysé |
| Postcondition | Un message d’alerte est affiché à l’utilisateur sur l’interface |
| Scenario de déroulement | |
| 1. L’utilisateur capture le trafic réseau 2. Il analyse ce trafic 3. Le système lui renvoi des messages d’alertes | |

* **Déconnecter un utilisateur**

Tableau 9: documentation du cas d'utilisation "se Déconnecter'

|  |  |
| --- | --- |
| Cas d’utilisation | Déconnecter un utilisateur |
| Acteur | Administrateur |
| Préconditions | L’administrateur doit être connecté au système |
| Scenario de déroulement | |
| 1. L’utilisateur appuie sur le bouton de déconnexion 2. Il est déconnecté et redirigé sur la page de connexion | |

* + 1. **Diagramme de séquences**

Si le diagramme de cas d’utilisation décrit les interactions des acteurs avec le système, lors de ces interactions les acteurs produisent des messages qui affectent le système informatique et appelle généralement des réponses de celui-ci. Le diagramme de séquence permet d’isoler et de représenter graphiquement ces messages [25].

Nous allons représenter les diagrammes de séquences d’un scenario de chaque cas d’utilisation présenté précédemment.

1. **Diagramme de séquence pour le cas d’utilisation « créer Compte »**

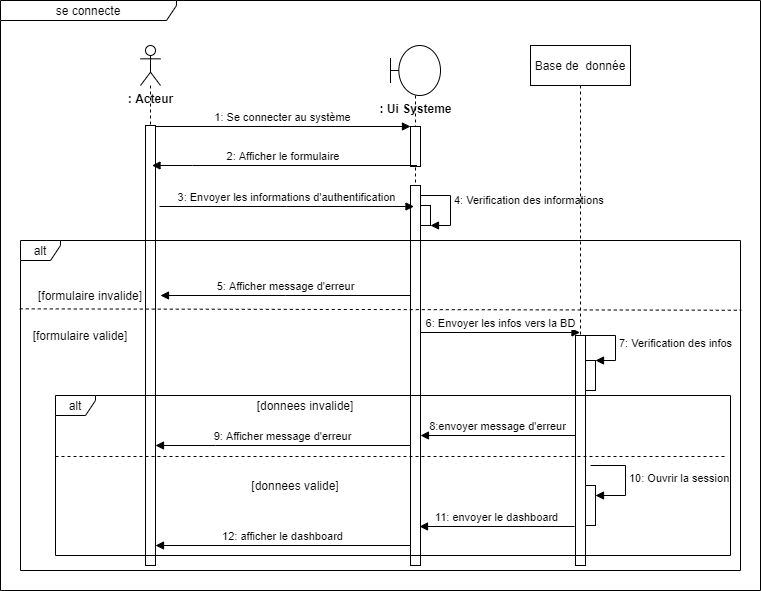


Figure 2‑3: Diagramme de séquence pour le cas d'utilisation "créerCompte"

1. **Diagramme de séquence pour le cas d’utilisation « se Connecter »**

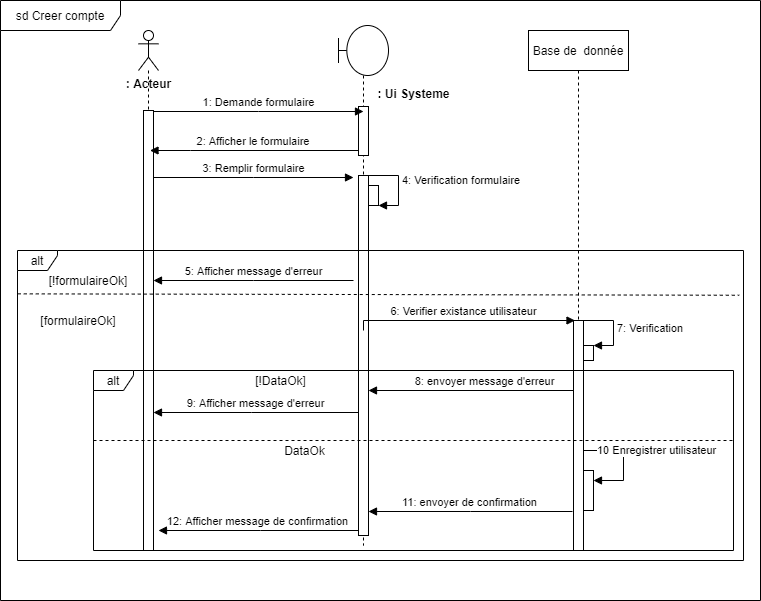


Figure 2‑4: Diagramme de séquence pour le cas d'utilisation "se Connecter"

1. **Diagramme de séquence pour le cas d’utilisation « configurer Système »**

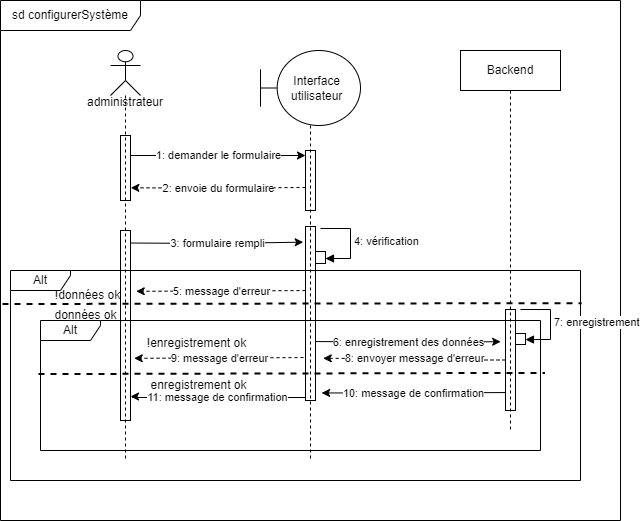
****

Figure 2‑5: diagramme de séquence pour le cas d'utilisation "configurer Système"

1. **Diagramme de séquence pour le cas d’utilisation « recolterTrafic »**

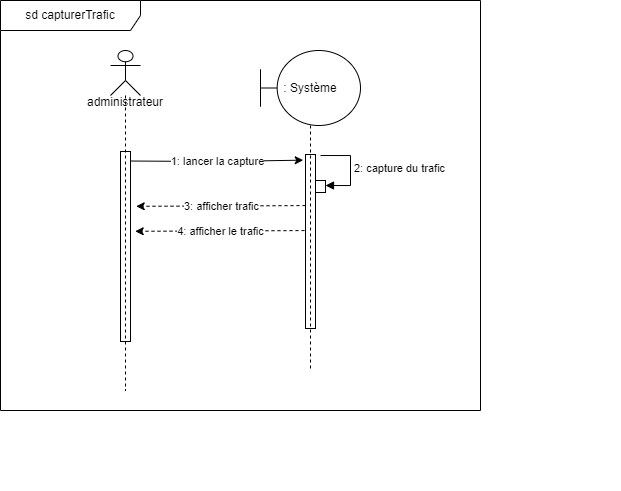
****

Figure 2‑6: diagramme de séquence pour le cas d'utilisation "capturer Trafic"

1. **Diagramme de séquence pour le cas d’utilisation « analyserTrafic »**

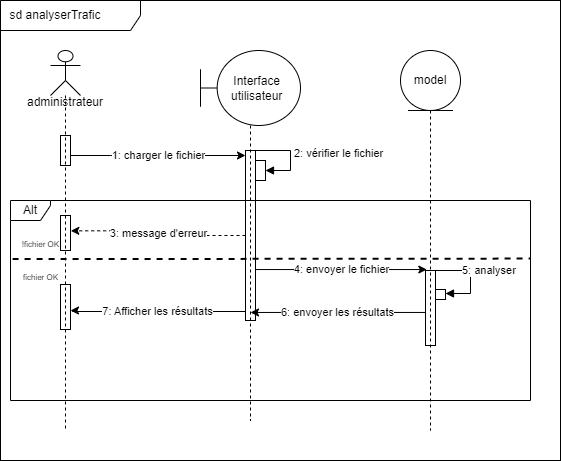
****

Figure 2‑7: Diagramme de séquence pour le cas d'utilisation "analyserTrafic"

1. **Diagramme de séquence pour le cas d’utilisation « alerter »**

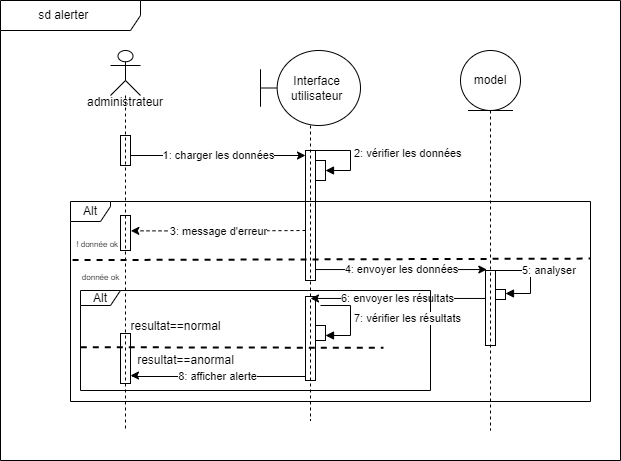
****

Figure 2‑8: diagramme de séquence pour le cas d'utilisation "alerter"

1. **Diagramme de cas d’utilisation pour le cas d’utilisation « seDéconnecter »**

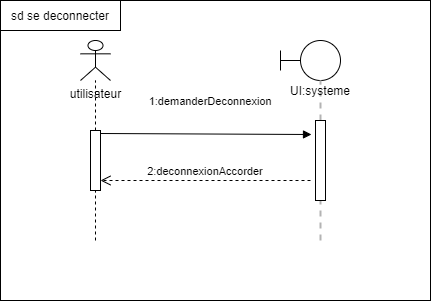


Figure 2‑9: diagramme de séquence pour le cas d'utilisation "seDeconnecter"

* 1. **Présentation des interfaces**

L’utilisateur interagit avec notre système grâce à des interfaces utilisateurs. Ceux-ci sont conçus dans le respect des règles d’IHM et lui permettent de réaliser les interactions attendues.

La conception d’une interface utilisateur doit respecter certaines règles d’ergonomie afin que la prise en main du système par l’utilisateur soit rapide et qu’il soit en mesure d’utiliser le système à son plein potentiel [25].

Dans cette section, nous présenterons les maquettes des interfaces utilisateurs respectant les règles de conception des interfaces hommes-machine.

1. **Interface de création de compte**

C’est par cette interface que l’utilisateur entre ses identifiants pour créer un compte, une fois toutes les données entrées il valide. Si les identifiants sont incorrects ou si elles correspondent déjà à un autre compte un message d’erreur est envoyé à l’utilisateur afin de lui faire savoir que la validation de son formulaire n’a pas été effective.

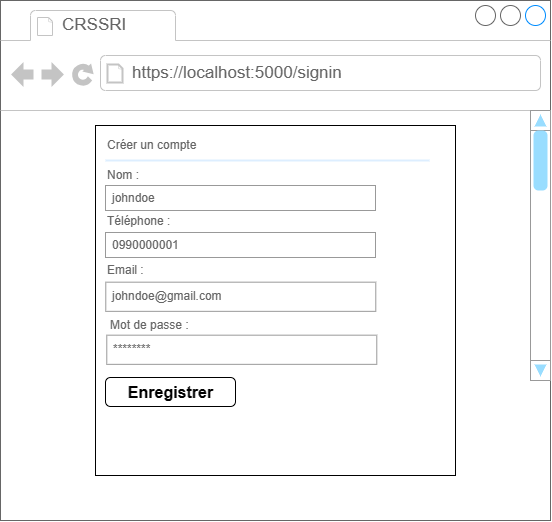


Figure 2‑10: maquette de l'interface de création de compte utilisateur

1. **Interface de connexion**

Après la création de compte, l’utilisateur est renvoyé vers cette page afin de se connecter au système. Là il entre les identifiants correspondants à ceux entré lors de la création de son compte utilisateur. Si ceux-ci ne correspondent pas alors, il s’affiche un message clair et concis lui expliquant l’erreur. Une fois ses données validées, il est redirigé vers une autre page.

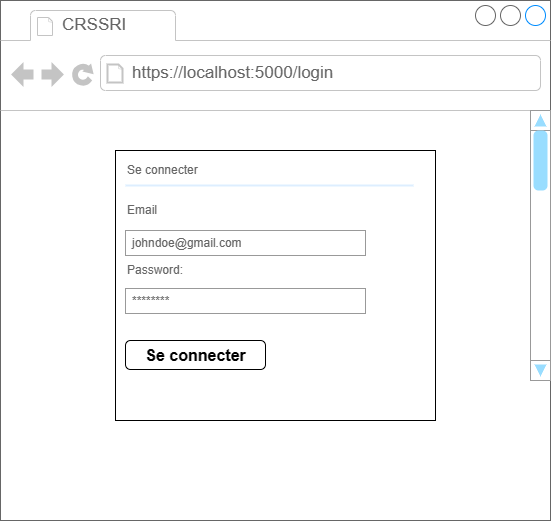


Figure 2‑11: maquette de l'interface de connexion de l’utilisateur

1. **Interface de global du système**

Une fois l’utilisateur connecté au système il accède à une interface comme celle de la figure 17. Cette maquette est un Template de toutes les interfaces que l’utilisateur pourra voir par la suite en choisissant la page à afficher dans le menu à gauche. Seul le contenu du grand boc à droite changera lors des interactions de l’utilisateur avec le système.

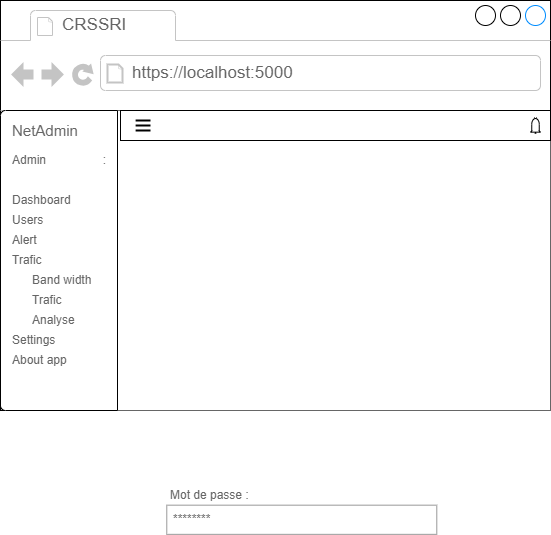


Figure 2‑12: maquette de l'interface globale du système

**Conclusion partielle**

Dans ce chapitre, nous avons présenté la conception du système de monitoring du réseau informatique. Nous avons présenté les diagrammes de conception de notre système entre autres le diagramme de cas d’utilisations et les diagrammes de séquences ainsi que les maquettes des interfaces. Notre système est un système interactionnel et n’a qu’un seul acteur notamment l’administrateur du réseau informatique.

Dans le prochain chapitre nous allons développer le processus d’implémentation de notre système, les technologies utilisées ainsi que la présentation des résultats.

# Implémentation et présentation des résultats

## Introduction

Dans ce dernier chapitre, nous allons présenter la réalisation de notre système de surveillance du trafic d’un réseau informatique en utilisant le machine Learning.

Dans le chapitre précédent nous avons modélisé notre système, dans ce présent nous allons présenter toutes les étapes par lesquelles nous sommes passé afin d’implémenter le système.

Nous avons conçu un modèle de classification qui permet de prédire sur base du trafic sur un réseau informatique si ce dernier est normal ou anormal.

## Présentation du dataset

Pour résoudre un probleme de détection d’anomalie en utilisant le machine Learning, il est impérieux d’avoir des données sur lesquels le model sera entrainé et par la suite testé. Ainsi, nous avons opté dans le cadre de notre travail d’utiliser le dataset NSL KDD. C’est un dataset qui est couramment utilisé particulièrement pour les probleme de détection d’intrusion sur un réseau informatique. Il est fiable parce qu’il présente moins des données redondantes que d’autres dataset tel que le KDD 99 qui autrefois était le plus utilisé pour cette question. En fait le NSL KDD dataset est une version corrigée du KDD 99 [22].

C’est un dataset avec des données etiquettées, il est composé de 42 colonnes, la 42ième colonnes est l’étiquette : soit normal ou anormal.

Les colonnes du datasets sont :

Tableau 10 : table des attributs du NSL-KDD dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Attributs du dataset | | | | | |
| *'duration'* | *'wrong\_fragment'* | *'su\_attempted'* | *'is\_guest\_login'* | *'same\_srv\_rate'* | *'dst\_host\_same\_src\_port\_rate'* |
| *'protocol\_type'* | *'urgent'* | *'num\_root'* | *'count'* | *'diff\_srv\_rate'* | *'dst\_host\_srv\_diff\_host\_rate'* |
| *'service'* | *'hot'* | *'num\_file\_creations'* | *'srv\_count'* | *'srv\_diff\_host\_rate'* | *'dst\_host\_serror\_rate'* |
| *'flag'* | *'num\_failed\_logins'* | *'num\_shells'* | *'serror\_rate'* | *'dst\_host\_count'* | *'dst\_host\_srv\_serror\_rate'* |
| *'src\_bytes'* | *'logged\_in'* | *'num\_access\_files'* | *'srv\_serror\_rate'* | *'dst\_host\_srv\_count'* | *'dst\_host\_rerror\_rate'* |
| *'dst\_bytes'* | *'num\_compromised'* | *'num\_outbound\_cmds'* | *'rerror\_rate'* | *'dst\_host\_same\_srv\_rate'* | *'dst\_host\_srv\_rerror\_rate'* |
| *'land'* | *'root\_shell'* | *'is\_host\_login'* | *'srv\_rerror\_rate'* | *'dst\_host\_diff\_srv\_rate'* | *'class'* |

## Prétraitement des données

Le dataset que nous avons utilisé a l’avantage d’offrir des données avec peu de disparité, ainsi nous n’avons pas eu besoin de réaliser énormément de prétraitement sur les données.

Les figures suivantes présentent la distribution des données dans le dataset.

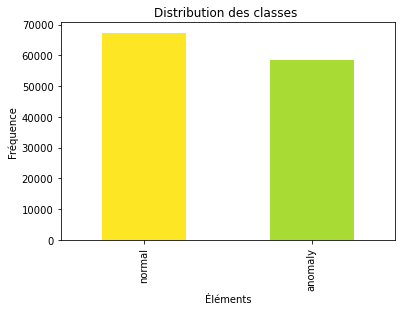


Figure 16 : Distribution des classes dans le dataset

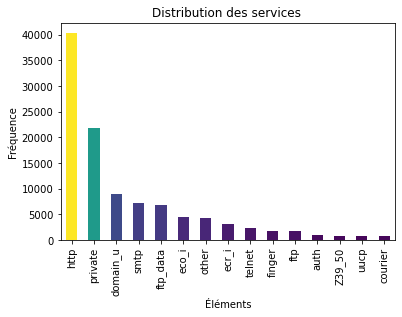


Figure 17: Distribution des services dans le dataset

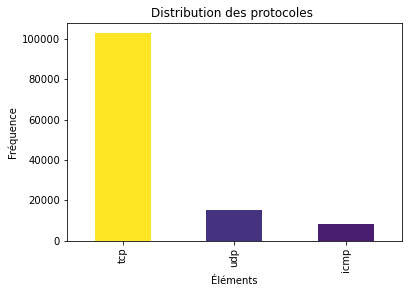


Figure 18: Distribution des protocoles dans le dataset

Le dataset est constitué des données bien reparties comme illustré sur la figure 16. Pour commencer, nous avons vérifié qu’il n’y a pas de doublons dans les données et les supprimer ; nous avons également appliquer une transformation sur les données car le dataset est composé de trois type de données entre autres le type Object, Integer et Float. Cette disparité nous a poussé à appliquer une transformation sur les données de type objet. Pour ce faire nous avons utilisé une fonction ***fit\_transform*** du module ***label\_encoder***.

La figure 19 et la figure 20 suivante présente les données avant et après transformation.

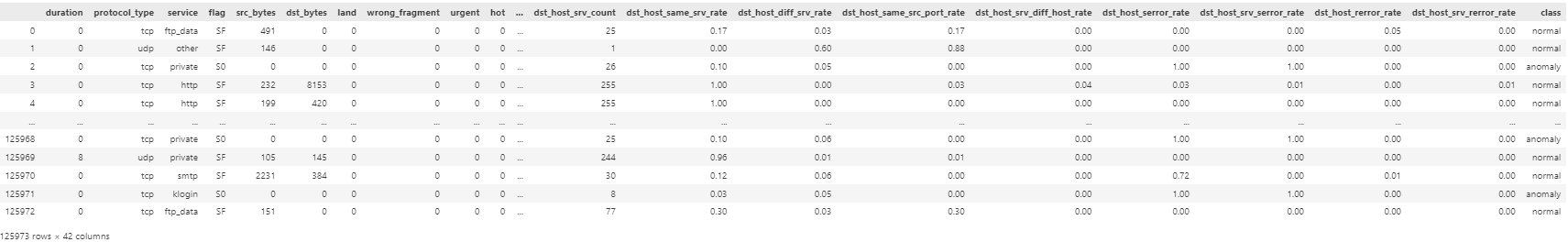


Figure 19: données du dataset avant la transformation

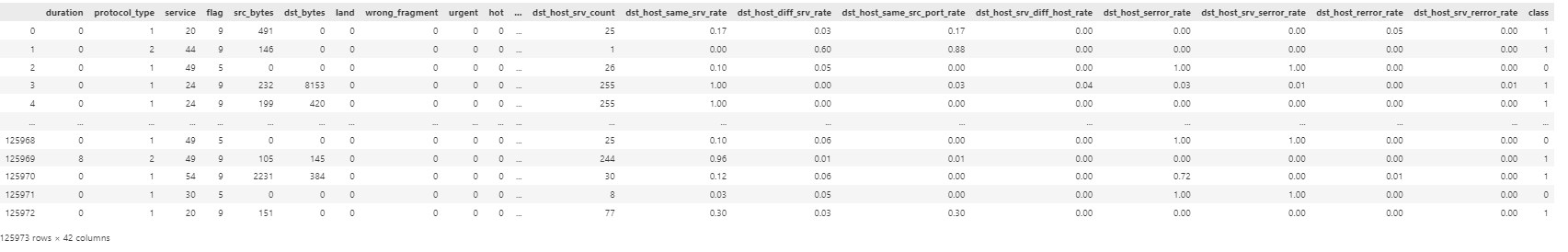


Figure 20: données du dataset après la transformation

### Feature selection

Une étude menée par … présente les 42 features qui constituent le dataset et leur impact dans la detection d’intrusion. En effet ils demontrent que pour une detection d’intrusion on a un taux de prediction maximal lorsqu’on utilise une combinaison des features de base, ceux relatif au hote et relatif au reseau. Ou des features de base et ceux relatif au hote.

En se basant sur les resultats presenté dans cette etude nous avons choisi d’entrainer…

La selection des caractères est un processus clé de notre méthodologie car pour des probleme de detection d’intrusion ou d’anomalies plusieurs element entre en jeu.

Dans la revue NSL DAtaset analysis une etude a été menée sur les colones constituant le dataset, il a été démontré que les meilleurs résultats sont obtenus soit en combisants les caractères basique (protocol\_type, service,…) avec les caractères relatif au hote ou soit en combinant les caractères basique avec les caractères relatif au reseau.

Ainsi nous avons choisi des caractères telques :

## Entrainement du model

Afin de faire un choix optimal de l’algorithme que nous utiliserons pour notre modèle nous avons entrainé le modèle en utilisant quatre différents algorithmes d’apprentissage supervisé par la suite comparer leur score de prédiction, la matrice de confusion ainsi que le temps de prédiction. Ces algorithmes sont : *Logistic Regression*, *Random forest*, *Decision tree* et *K Nearest Neighbors*.

Les score de prédiction et le temps de prédiction de chaque algorithme est presente dans le tableau suivant :

Tableau 11: table comparative des performances des quatre algorithmmes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Algorithmes de classification | | | |
| ***Logistic Regression*** | ***Random forest*** | ***Decision tree*** | ***K Nearest Neighbors*** |
| Score de prédiction | 0. 927313 |  | 0. 997592 | 0.986029 |
| Temps de pédiction |  |  |  |  |

Les matrices de confusions sont présentées dans les figures suivantes :

De ce qui precede, nous remarquons que l’algorithme Decision tree offre le meilleur score de prediction mais son temps de prediction est faible comparativement à celui de Logistic regression, Les performances attendues de notre système sont : une prediction avec le moins d’erreur afin de réduire les predictions fausse negative et fausses positives et temps réel.

L’algorithme qui maximise ces performances est donc le Decision tree.

## Presentation des résultats

Notre modele concu avec le machine learning basé sur l’algorithme de classification decision tree permet de prédire si un traffic sur un reseau est normal ou anormal.

Un traffic normal est… tandisqu’un anormal est…

Nous avons testé notre model avec le donnees de test du dataset NSL KDD et sur un traffic réel.

Les résultats obtenus sont presente dans le tableau 12 suivante :

Tableau 12: Tableau des résultats pour les quatre algorithmes

|  |  |  |  |
| --- | --- | --- | --- |
| Modèle | Score d’entrainement | Score de test | Temps de prédiction pour une donnée |
| Logistic Regression | 0.928998 | 0.927313 |  |
| Random forest |  |  |  |
| Decision Tree | 0.999932 | 0.997592 |  |
| KNN | 0.98731 | 0.986029 |  |

## Déploiement du modèle

Une fois que nous avons testé notre modèle sur un trafic réel nous avons obtenu un score de 0.7781227821149752

Pourcentage de prédictions normales : 52.32%

Pourcentage de prédictions anormales : 47.68%

Les matrices de confusion

Nous rema

## Implementation de l’interface d’administration du réseau

Nous avons implementé cette interface en utilisant le framework web python Flask.

Dans cette section nous allons présenter les interfaces de notre système.

Interface de creation de compte

Interface de connexion

Interface d’affichage des utilisateurs réseaux

Interface de capture et analyse de trafic reseau

Interface du dashboard

Conclusion partielle

Ce chapitre était consacré à l’implémentation de notre système, nous avons présentés toutes les étapes par lesquelles nous sommes passées pour la réalisation de notre système de surveillance de réseau informatique en utilisant le machine Learning ainsi que les résultats obtenus par note système.

Conclusion générale

La conception d’un système de surveillance de reseau informatique est le but poursivi dans ce travail. Nous avons commencé par une introduction dans laquelle nous avons présenter l’idee qui nous a guidé tout au long de ce travail notamment la problématique qui était de concevoir un système pouvant permettre à un administrateur de reseau informatique de surveiller efficacement son reseau. La solution que nous avons apporté à cette problématique par notre travail est ce mettre en place un système qui permet de réaliser cette surveillance en offrant une interface par laquelle on peut capturer le traffic sur le reseau et analyser ce dernier en le passant dans un model qui permet de dire si ce trafic est nomal ou sinon générer une alerte dans le système.

Dans les chapitre un nous avons parlé brièvement des generalites sur les réseaux informatiques et leur sécurité, element important pour comprendre la suite du travail

Dans le chapitre deux nous avons presenté la conception de notre systemee en présentant les specifications ainsi que les diagrammes de sequences de notre système.

Enfin dans le chapitre trois nous avons parlé de l’implementation et la presentation des résultats de notre système.

Ceci nous permet de conclure que l’utilisation du machine learning dans le domaine de surveillance des réseaux informatiques presente d’enorme potentiel car ca offre des résultats meilleurs que les techniques basiques.

Egalement le choix de l’algorithme est un facteur capital dans la réalisation d’un model de surveillance de reseau informatique

On peut affirmer que pour une meilleure surveillance il est important de focaliser son attentions sur les éléments capital qui permettent de mieux prédire si un trafic est normal ou anormal et réduire le temps perdu.

Au bout de ce travail nous avons un système fonctionnel et efficace qui permet de surveiller un reseau informatique en temps relle.

Bibliographie

|  |  |
| --- | --- |
| [1] | A. L. R. F. a. N. C. Chuck Swanson, Security+ - A compTIA Certification Student Manual, Rochester: Element K, 2003. |
| [2] | I. G. a. a. Ansam Khraisat, «Survey of Intrusion Detection Systems : techniques, datasets and challenges,» *Open Acess,* 2019. |
| [3] | D. Mach, Computer Networking for beginners. |
| [4] | G. Pujolle, Initiation aux réseaux : Cours et exercices, Paris: Eyrolles, 2001. |
| [5] | A. Géron, Wifi professionnel : la norme 802.11 le deploiement, la sécurité, Paris: Dunod, 2009. |
| [6] | J.-L. Montagnier, Réseaux d'entreprises par la pratique, Eyrolles. |
| [7] | D. S. Danièle DROMARD, Architecture des réseaux, Paris: PEARSON Education, 2009. |
| [8] | D. Minoli, Hotspot Networks, McGRAW-HILL TELECOMMUNICATIONS, 2003. |
| [9] | «Cisco,» [En ligne]. Available: https://www.cisco.com/. [Accès le 21 mars 2023]. |
| [10] | [En ligne]. Available: https://www.kaspersky.fr/resource-center/definitions/what-is-an-ssid. |
| [11] | [En ligne]. Available: https://www.malekal.com/quest-ce-que-la-bande-passante-debit-wifi-et-dans-le-hardware/#Qu8217est-ce\_que\_la\_bande\_passante\_debit\_reseau\_et\_internet. |
| [12] | C. Hunt, TCP/IP Network Administration, Tokyo: O'REILLY, 2002. |
| [13] | B. A. Forouzan, TCP/IP Protocol suite, The McGraw-Hill Companies, 2010. |
| [14] | J. Ross, Network Know-how : An essential guide for the accidental admin, San francisco: no starch press, 2009. |
| [15] | [En ligne]. Available: https://fr.wikipedia.org/wiki/User\_Datagram\_Protocol. |
| [16] | [En ligne]. Available: https://www.malekal.com/udp-user-datagram-protocol-definition-structure-du-protocole/. |
| [17] | [En ligne]. Available: https://www.malekal.com/udp-user-datagram-protocol-definition-structure-du-protocole/. |
| [18] | [En ligne]. Available: https://www.cloudflare.com/fr-fr/learning/ddos/glossary/internet-control-message-protocol-icmp/. |
| [19] | [En ligne]. Available: https://www.cisco.com/c/fr\_ca/support/docs/smb/switches/cisco-small-business-300-series-managed-switches/smb1100-internet-control-message-protocol-icmp-filtering-configurati.html. |
| [20] | [En ligne]. Available: https://blogs.manageengine.com/fr/2022/02/18/solution-de-surveillance-snmp-des-informations-importantes-sur-la-surveillance-en-un-clic.html. |
| [21] | M. G. Naugle, Illustrated TCP/IP, Wiley Computer Publishing, John Wiley & Sons, Inc. , 1998. |
| [22] | A. MOUHLI, Administration système réseaux. |
| [23] | H. Gomaa, Software modeling and design, New York: Cambrifge university Press, 2011. |
| [24] | «Medium,» [En ligne]. Available: https://towardsdatascience.com/a-deeper-dive-into-the-nsl-kdd-data-set-15c753364657. |
| [25] | P. Roques, Les cahiers du programmeur : UML 2, modeliser une application web, Paris: Eyrolles. |
| [26] | M. I. N. V. Patrick, *slides de cours d'IHM : Interface Homme Machine,* 2022-2023. |
| [27] | JUnit.org. [En ligne]. Available: http://www.junit.org. [Accès le 2 Janvier 2019]. |
| [28] | L. B. William Stallings, Computer security: Principles and Practice, New York: Pearson Education, 2018. |
| [29] | S. K. S. Pretti Aggarwal, "Analysis of KDD Dataset Attributes - Class wise For Intrusion," in *3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015)*, India, 2015. |
| [30] | C. W. Laurent Bloch, Sécurité informatique : Principes et méthodes, Paris: Eyrolles, 2007. |
| [31] | S. Ghernaouti, Cybersécurité : Analyser les risques et mettre en oeuvre des solutions, 11, rue Paul Bert 92240 Malakoff: Dunod, 2019. |
| [32] | R. Anderson, Security engineering : A guide to building dependable distributed systems, Wiley, 2013. |
| [33] | N. J. JACK CARAVELLI, Cyber Security : Threat and response for Governement and Business, California: Praeger, 2019. |
| [34] | M. Harrison, Machine Learning: les fondamentaux, France: First, 2020. |
| [35] | G. Saint-Cirgue, Apprendre le Machine Learning en une semaine, 2019. |
| [36] | A. R. A. T. Mehryar Mohri, Foundations of Machine Learning, USA: Massachusetts Institute of Technologies, 2018. |
| [37] | A. C. Faul, A concise Introduction to Machine Learning, Taylor & Francis Group, 2020 . |
| [38] | J. Brownlee, Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch, 2016. |
| [39] | M. E. A. C. T. Kwangjo Kim, Network Intrusion Detection using Deep Learning : A future learning aproach, Singapore: Springer, 2018. |
| [40] | J. J. P. Zhenwei Yu, Intrusion detection : A machine Learning Approach - Vol.3, Londres: Imperial College Press, 2011. |
| [41] | P. Singh, Deploy Machine Learning Models to Production, Bangalore, Karnataka, India: Apress, 2021. |
| [42] | J. W. a. S. Sims, Python, scapy and fuzzing, SANS Institute, 2019. |
| [43] | A. A. Patel, Hands-on Unsupervised Learning using Python : How to Build Applied Machine Learning Solutions from Unlabelled data, USA: O’Reilly, 2019. |