REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTÈRE DE l'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ **S**AAD **D**AHLAB **B**LIDA 1

Faculté des Sciences

Département : Informatique

Mémoire de Fin d'Étude pour l'Obtention du Diplôme de Master en Informatique

OPTION : Sécurité des Systèmes d'Information

Thème :

Développement d'un Système d'Accès Intelligent En Utilisant les Méthodes d'Apprentissage Automatique

---

Organisme d'accueil : Ministère de la Poste, des Télécommunications, des Technologies et du Numérique.

Réalisé par :

BOUGHELIT Meriem

MOUSSOUS Soumia

Proposé par : Mme GHEBGHOUB Yasmine

Mr CHERIF ZAHAR Amine	**Président**
Mme BACHA Sihem	**Examinatrice**
Mme GHEBGHOUB Yasmine	**Promotrice**
Mr BEKHTI Hamza	**Encadreur**

Promotion : 2019-2020

Remerciements

Nos premiers remerciements vont à ALLAH le tout Puissant qui nous a donné la force, la santé et la volonté pour réaliser ce modeste travail.

Nous remercions notre promotrice Mme GHEBGHOUB Yasmine pour la confiance qu'elle nous a accordée en nous proposant ce travail, pour sa participation, sa disponibilité, son soutien et de nous avoir guidés dans la réalisation de ce travail.

Nous tenons également à remercier notre encadreur Mr BEKHTI Hamza pour avoir accepté de diriger ce travail et pour les précieuses remarques qu'il nous a apportées pour faire des améliorations sur ce travail et l'appliquer au sein de sa sous-direction.

Nous remercions vivement les membres de jury pour nous avoir fait l'honneur d'accepter d'examiner notre travail.

Sans oublier nos très chères familles et nos amis surtout pour leurs encouragements et leur soutien moral, merci à toutes et à tous.

Résumé

La sécurité des systèmes d'information est une problématique d'une importance majeure pour les individus ainsi que pour les entreprises. Elle repose sur la mise en place d'une politique de sécurité autour de ces systèmes. Pour compléter cette politique de sécurité, il est nécessaire d'avoir des systèmes de contrôle d'accès qui permettent d'assurer la protection contre le vol, la perturbation et la modification non autorisé de l'information. Aujourd'hui, le système d'accès intelligent est devenu le moyen idéal pour gérer les accès de tous les utilisateurs, tout en garantissant la sécurité de leurs données. Il utilise des technologies avancées pour améliorer encore la précision et la sécurité du processus de vérification de droit d'accès.

Dans notre travail nous proposons un système de sécurité intelligent basé sur le principe de modèle de contrôle d'accès (ORBAC). Ce modèle vise à donner aux utilisateurs la possibilité de contrôler la sécurité de leurs données. Il offre la faculté d'exprimer des permissions, des obligations, des interdictions et même des recommandations dépendant bien évidemment des contextes. En effet notre système intelligent utilisera aussi le principe de l'apprentissage supervisé. Cette proposition permettra de réduire l'intervention de l'homme sur le contrôle d'accès et augmentera la confidentialité et l'intégrité des données. Nous avons implémenté notre solution en utilisant un algorithme qui permet de classer les utilisateurs (Accès ou Non accès). Cette solution ne demande pas une base de données ou une table des utilisateurs, car notre système va se baser sur les poids des requêtes. Il prend en charge une base de connaissance qui contient les permissions de tous les utilisateurs.

Mots clés : contrôle d'accès, modèle de contrôle accès (ORBAC), système d'accès intelligent, apprentissage supervisé, poids des requêtes, base de connaissance, permission.

Abstract

The security of information systems is a problem of major importance for individuals as well as for companies. It is based on the establishment of a security policy around these systems. To complete this security policy, it is necessary to have acces control systems that ensure protection against theft, disruption and unauthorized modification of information. Today, the intelligent access system has become the ideal way to manage access for all users, while ensuring the security of their data. It uses advanced technologies to further improve the accuracy and security of the access right verification process.

In our work we propose an intelligent security system based on the principle of access control model (ORBAC). This model aims to give users the possibility to control the security of their data. It offers the ability to express permissions, obligations, prohibitions and even recommendations, obviously depending on the context. Indeed our intelligent system also uses the principle of supervised learning. This proposition will reduce human intervention in access control and increase the confidentiality and integrity of data. We implemented our solution using an algorithm that classifies users (Access or Not access). This solution does not require a database or a user table, because our system will be based on the weights of the queries. It supports a knowledge base which contains the permissions of all users.

Keywords : access control, access control model (ORBAC), intelligent access system, supervised learning, weights of the queries, knowledge base, permission.

Table des matières

Introduction générale

1. Contexte 1
2. Problématique 1
3. Objectif 1
4. Organisation du mémoire 2

Chapitre I : La sécurité informatique et les techniques de contrôle d'accès

I.1 Introduction 4
I.2 La sécurité informatique 4
I.2.1 Définition de la sécurité informatique 4
I.2.2 Objectifs de la sécurité informatique 4
I.2.2.1 La confidentialité 4
I.2.2.2 L'authentification 4
I.2.2.3 L'intégrité 4
I.2.2.4 La disponibilité 4
I.2.2.5 La non-répudiation 4
I.2.3 Les attaques 5
I.2.3.1 C'est quoi une attaque 5
I.2.3.2 IP Spoofing 5
I.2.3.3 ARP Spoofing 5
I.2.3.4 Sniffing 5
I.2.3.5 Déni de service 5
I.2.3.6 Man in the middle 5
I.2.3.7 Les injections SQL 6
I.2.4 Les logiciels malveillants 6
I.2.4.1 Les Virus 6
I.2.4.2 Les Vers 6
I.2.4.3 Les chevaux de Troie 6
I.2.4.4 Le logiciel espion 6
I.2.5 Techniques de défense et de sécurité 6
I.2.5.1 Authentification 7
I.2.5.2 Sensibilisation du personnel 7
I.2.5.3 Firewall (pare-feu) 7
I.2.5.4 Antivirus 7
I.2.5.5 VPN 7

I.2.5.6 IDS 8
I.2.5.7 IPS 8
I.2.5.8 Serveur proxy 8
I.2.5.9 Cryptographie 8
I.2.5.10 Contrôle d'accès 9
I.3 Le contrôle d'accès 9
I.3.1 Définition de contrôle d'accès 9
I.3.2 Définition d'une politique de contrôle d'accès 9
I.3.3 Les modèles de contrôle d'accès 10
I.3.3.1 DAC (Discretionary Access Control) 10
I.3.3.2 MAC (Mandatory Access Control) 11
I.3.3.3 RBAC (Role-Based Access Control) 12
I.3.3.4 Modèle de contrôle d'accès à basé d'équipe (TMAC) 12
I.3.3.5 ORBAC (Organization Based Access Control) 13
I.3.4 Comparaison entre les différents modèles de contrôle d'accès 18
I.4 Conclusion 18
Chapitre II : L'apprentissage automatique
II.1 Introduction 20
II.2 Définition de l'intelligence artificielle (IA) 20
II.3 Définition d'un système intelligent 20
II.4 Définition de l'apprentissage automatique 21
II.5 Les principales tâches de l'apprentissage automatique 21
II.6 Les types d'apprentissage 22
II.6.1 L'apprentissage supervisé 22
II.6.2 L'apprentissage non supervisé 23
II.7 Les algorithmes de classification 24
II.7.1 L'algorithme k-Means (Macqueen, 1967) 24
II.7.2 L'algorithme SVM (Support Vector Machine) 26
II.7.3 L'algorithme KNN 27
II.8 Conclusion 29
Chapitre III : Modélisation et Conception
III.1 Introduction 31
III.2 La méthode EBIOS (Expression des Besoins et Identification des Objectifs de Sécurité) 31
III.2.1 Présentation de la démarche EBIOS 31
III.2.2 Module 1 : Étude du contexte 32
III.2.3 Module 2 : Étude des événements redoutés 40
III.2.4 Module 3 : Étude des scénarios de menaces 41

III.2.5 Module 4 : Étude des risques 42
III.2.6 Module 5 : Étude des mesures de sécurité 43
III.3 L'approche UML (*Unified Modeling Language*) 44
III.3.1 Définition 44
III.3.2 Diagramme de cas d'utilisation 44
III.3.3 Diagramme de séquence 45
III.3.3.1 S'authentifier 45
III.3.3.2 Demander l'accès à la ressource 46
III.3.3.3 Faire une activité sur la ressource 47
III.3.4 Diagramme de classe 48
III.4 Conclusion 48
Chapitre IV : Implémentation
IV.1 Introduction 50
IV.2 Implémentation du système 50
IV.2.1 Description de fonctionnement du système 50
IV.2.2 Algorithme proposé basé sur l'apprentissage supervisé 52
IV.2.3 Exemple des tests 54
IV.2.4 La cryptographie 55
IV.3 Outils de développement 56
IV.3.1 Outils matériels 56
IV.3.2 Outils logiciels 56
IV.3.3 Langages de programmation 57
IV.4 Présentation de l'application 59
IV.4.1 Espace administrateur 60
IV.4.2 Espace utilisateur 63
IV.5 Conclusion 65
Conclusion générale 66
Bibliographie 67

Liste des figures

Figure 1 : Matrice de contrôle d'accès [9] 11
Figure 2 : Exemple de contrôle d'accès avec les niveaux de sécurité [1] 11
Figure 3 : Attribution des permissions en RBAC [1] 12
Figure 4 : Modèle ORBAC [13] 13
Figure 5 : La relation Habilite [17] 14
Figure 6 : La relation Utilise [17] 15
Figure 7 : La relation Considère [17] 16
Figure 8 : La relation Définit [17] 17
Figure 9 : Structure du modèle OrBAC [12] 17
Figure 10 : Apprentissage supervisé 23
Figure 11 : Exemple sur l'algorithme K-Means [37] 25
Figure 12 : Problème de classification à deux classes avec une séparatrice linéaire et non linéaire [41] 26
Figure 13 : Exemple d'algorithme KNN [42] 28
Figure 14 : Démarche EBIOS [44] 32
Figure 15 : Organigramme de l'entreprise 33
Figure 16 : Architecture du réseau de l'entreprise 35
Figure 17 : Diagramme de cas d'utilisation global 44
Figure 18 : Diagramme de séquence « S'authentifier » 45
Figure 19 : Diagramme de séquence « Demander l'accès à la ressource » 46
Figure 20 : Diagramme de séquence « Faire une activité sur la ressource » 47
Figure 21 : Diagramme de classe 48
Figure 22 : Architecture du système 50
Figure 23 : Les permissions des utilisateurs 51
Figure 24 : Les permissions des utilisateurs pour le contexte « urgence » 52
Figure 25 : La fonction de cryptage/décryptage de l'algorithme AES 55
Figure 26 : Le résultat d'utilisation de fonction de cryptage pour le mot de passe 55
Figure 27 : L'interface principale de l'application 59
Figure 28 : Page d'authentification 59
Figure 29 : L'interface principale de « Espace administrateur » 60
Figure 30 : Gestion d'utilisateur 60
Figure 31 : Ajout d'un utilisateur 61
Figure 32 : Modification d'un utilisateur 61
Figure 33 : Suppression d'un utilisateur 62
Figure 34 : Détails d'un utilisateur 62
Figure 35 : Formulaire d'accès 63
Figure 36 : Message d'erreur pour l'organisation 64
Figure 37 : Exemple d'une ressource 64
Figure 38 : Vérification de contexte 65
Figure 39 : Message de refus d'accès 65

Liste des tableaux

Tableau 1 : Comparaison entre les modèles de contrôle d'accès.....18
Tableau 2 : Présentation des sources de menaces.....37
Tableau 3 : Présentation des biens support à protéger.....38
Tableau 4 : Présentation des échelles de disponibilité, d'intégrité et de confidentialité [45].....39
Tableau 5 : Échelle de gravité [47].....39
Tableau 6 : Échelle de vraisemblance [47].....39
Tableau 7: Mesures de sécurité existantes.....40
Tableau 8 : Étude des événements redoutés.....41
Tableau 9 : Étude des scénarios de menaces.....42
Tableau 10 : Descriptions des cas d'utilisation du diagramme de cas d'utilisation global.....45
Tableau 11 : Exemple des tests.....54
Tableau 12 : Exemple des tests après l'utilisation de contexte.....54

Liste des schémas

Schéma 1 : Direction générale de la société de l'information.....33

Introduction générale

1. Contexte

La sécurité est un sujet qui touche tous les composants du système d'information, y compris l'environnement et les utilisateurs. De nos jours, diverses informations sont échangées entre les systèmes, ceci ne peut pas être réalisé sans des problèmes de sécurité. Pour cela chaque organisation doit avoir une politique de sécurité qui permet d'assurer la protection contre le vol, la divulgation et la modification de l'information. Cette politique sera en charge de définir les privilèges d'accès des utilisateurs aux informations ce qui rend le système plus sécurisé. Donc Il faut connaître les ressources de l'organisation à protéger et de maîtriser le contrôle d'accès et les droits des utilisateurs du système d'information. La sécurité ne peut être assurée à cent pour cent. Toutefois, l'apprentissage automatique a empellement donner un nouveau sang dans le développement des systèmes de sécurité. Il peut résoudre les problèmes auxquels le contrôle d'accès est confronté, en développant des méthodes permettant aux ordinateurs de traiter des tâches sans être explicitement programmés pour chacune. Cela permet d'améliorer les performances du système d'information.

2. Problématique

Aujourd'hui, la sécurité est un enjeu majeur pour les entreprises. Les menaces numériques ne cessent d'augmenter. Pour les contrer, les entreprises s'appuient sur des systèmes d'accès à l'information, afin de résoudre les problèmes liés à la confidentialité et l'intégrité des données des différents utilisateurs. Ces accès sont déterminés par l'intervention humaine, ce qui rend le système moins sécurisé. Mais avec le développement des mesures de sécurité, on peut poser la question suivante : Comment réduire l'intervention de l'homme pour que le système d'accès gère lui-même ?

3. Objectif

L'objectif principale de notre travail c'est faire la conception et le développement d'un système d'accès intelligent. Notre étude a ciblé la protection de l'information et les accès dans l'organisation en générale, et ses différentes structures en particulier, dont les permissions des utilisateurs sont déterminées par le système. Ainsi, notre système d'accès doit traiter d'une manière automatique et intelligente toutes les demandes d'accès aux ressources.

4. Organisation du mémoire

Notre mémoire comporte quatre chapitres :

- Dans le premier chapitre nous avons parlé de la sécurité informatique et leurs principaux concepts. Ensuite, nous avons présenté les techniques de contrôle d'accès ainsi que les différents modèles de contrôle d'accès, et une comparaison entre ces modèles.
- Dans le deuxième chapitre, nous allons présenter l'apprentissage automatique, ses types ainsi que quelques algorithmes les plus utilisées.
- Pour le troisième chapitre, il est dédié à la modélisation en utilisant la méthode EBIOS pour faire une analyse de risques au sein de l'entreprise, ainsi qu'une représentation UML de notre travail.
- Le dernier chapitre, est consacré à l'implémentation de notre solution avec les différents outils utilisés.

Nous clôturons cette mémoire par une conclusion générale pour résumer notre travail et évaluer les résultats obtenus, ainsi que nos perspectives d'avenir pour améliorer la solution.

Chapitre I : La sécurité informatique et les techniques de contrôle d’accès

I.1 Introduction

De nos jours, la sécurité informatique se révèle une priorité de haute importance pour protéger le système d'information d'une organisation. Le contrôle d'accès présente un outil très important dans la sécurité des systèmes d'information. Il donne l'autorisation d'accéder aux ressources demandées (fichier, programme...). Il s'agit d'un ensemble des permissions et des interdictions pour assurer la confidentialité, l'intégrité et la disponibilité de l'information au sein de l'organisation.

Dans ce chapitre nous allons présenter la sécurité informatique ainsi que les techniques de contrôle d'accès.

I.2 La sécurité informatique

I.2.1 Définition de la sécurité informatique

Il existe plusieurs définitions de sécurité informatique parmi eux :

Définition 1 : La sécurité informatique est l'ensemble des moyens techniques, organisationnels, juridiques et humains nécessaires et mis en place pour conserver, rétablir, et garantir la sécurité des systèmes informatiques. Elle est intrinsèquement liée à la sécurité de l'information et des systèmes d'information. [1]

Définition 2 : La sécurité informatique est l'ensemble des moyens mise en œuvre pour réduire la vulnérabilité d'un système contre les menaces accidentelles ou intentionnelles. Elle s'occupe de la prévention d'actions non autorisées par les utilisateurs d'un système informatique, afin d'assurer certaines notions que nous allons définir dans ce qui suit. [2]

I.2.2 Objectifs de la sécurité informatique

On peut définir les objectifs de la sécurité informatique comme suit :

I.2.2.1 **La confidentialité** : les données ne doivent être accessible que par les personnes autorisées, c'est à dire le système empêche les utilisateurs non autorisés de voir ou lire une information confidentielle.

I.2.2.2 **L'authentification** : consiste à vérifier l'identité d'un utilisateur, c'est-à-dire garantir à chacun des correspondants, que son partenaire est bien celui qu'il croit être.

I.2.2.3 **L'intégrité** : il faut pouvoir garantir que les données protégées n'ont pas été modifiées ou altérées par une personne non autorisée.

I.2.2.4 **La disponibilité** : les données doivent restées accessibles aux utilisateurs à tout moment.

I.2.2.5 **La non-répudiation** : consiste à assurer que le message a bien été envoyé par un émetteur et reçu par un destinataire.

I.2.3 Les attaques

Les systèmes d'information sont aujourd'hui de plus en plus ouverts sur Internet. Cette ouverture, a priori bénéfique, pose néanmoins un problème majeur : il en découle un nombre croissant d'attaques.

I.2.3.1 C'est quoi une attaque ?

Une attaque c'est le résultat de l'exploitation d'une faille d'un système informatique (système d'exploitation, logiciel, erreur de configuration, ...etc.) à des fins non connue par l'exploitant du système et il est généralement répudiable [2]. Parmi les attaques les plus connues on a :

I.2.3.2 IP Spoofing

Le Spoofing consiste à se faire passer pour un autre système en falsifiant son adresse IP. Le pirate commence par choisir le système qu'il veut attaquer. Après avoir obtenu le maximum de détails sur ce système cible, il détermine les systèmes ou adresses IP autorisés à se connecter au système cible. Le pirate attaque ensuite le serveur cible en utilisant l'adresse IP falsifiée. [2]

I.2.3.3 ARP Spoofing

Le but de cette attaque est de rediriger le trafic d'une machine vers une autre. Grâce à cette redirection, une personne mal attentionnée peut se faire passer pour une autre. De plus, le pirate peut ré-router les paquets qu'il reçoit vers les véritables destinataires, ainsi l'utilisateur usurpé ne se rendra compte de rien. [2]

I.2.3.4 Sniffing

Le Sniffing ou reniflement de trafic constitue l'une des méthodes couramment utilisées par les pirates informatiques pour espionner le trafic sur le réseau. Dans la pratique, les hackers font généralement recours à ce procédé, pour détecter tous les messages circulant sur le réseau en récupérant des mots de passe et des données sensibles. [2]

I.2.3.5 Déni de service

Le déni de service est une attaque visant à rendre indisponible un service. Ceci peut s'effectuer de plusieurs manières : par le biais d'une surcharge réseau, rendant ainsi la machine totalement injoignable, ou bien de manière applicative en crashant l'application à distance. [3]

I.2.3.6 Man in the middle

L'attaque « man in the middle » littéralement « attaque de l'homme au milieu », est un scénario d'attaque dans lequel un pirate écoute une communication entre deux interlocuteurs et falsifiée éventuellement les échanges à leur insu. [2]

I.2.3.7 Les injections SQL

Tout comme les attaques de scripts, les injections SQL profitent de paramètres d'entrée non vérifiés. Comme leur nom l'indique, le but des injections SQL est d'injecter du code SQL dans une requête de base de données. Ainsi, il est possible de récupérer des informations se trouvant dans la base (exemple : des mots de passe) ou encore de détruire des données. [3]

I.2.4 Les logiciels malveillants

Ce sont des logiciels développés par des hackers dans le but de nuire à un système d'information. [4]

I.2.4.1 Les Virus

C'est un petit programme qui a la faculté de se reproduire automatiquement. Il va recopier son propre code tel quel, ou en le modifiant, dans des éléments qui sont déjà dans l'ordinateur. Le plus souvent son but est de nuire. [5]

I.2.4.2 Les Vers

Un ver est un programme autonome qui se reproduit et se propage à l'insu des utilisateurs. Contrairement aux virus, un ver n'a pas besoin d'un logiciel hôte pour se dupliquer. Le ver a habituellement un objectif malicieux. [4]

I.2.4.3 Les chevaux de Troie

Un cheval de Troie est une forme de logiciel malveillant déguisé en logiciel utile. Son but : se faire exécuter par l'utilisateur, ce qui lui permet de contrôler l'ordinateur et de s'en servir pour ses propres fins. Généralement d'autres logiciels malveillants seront installés sur votre ordinateur, tels que permettre la collecte frauduleuse, la falsification ou la destruction de données. [4]

I.2.4.4 Le logiciel espion

(Espiogiciel ou logiciel espion) est un programme ou un sous-programme, conçu dans le but de collecter des données personnelles sur ses utilisateurs et de les envoyer à son concepteur, ou à un tiers via Internet ou tout autre réseau informatique, sans avoir obtenu au préalable une autorisation explicite et éclairée desdits utilisateurs. [4]

I.2.5 Techniques de défense et de sécurité

La variété et la disponibilité des outils d'attaques augmentent le risque des intrusions. Par conséquent les administrateurs s'appuient sur diverses solutions dans le but de maintenir la protection du réseau informatique. Voici quelques solutions proposées : [5]

I.2.5.1 Authentification [6]

C'est vérifier la véracité des utilisateurs, du réseau et des documents. L'authentification n'est pas seulement l'utilisation d'une suite de mots de passe, mais il existe de nombreuses variétés de mécanismes possibles. On peut ranger ces mécanismes en trois catégories qui vérifient au moins l'un des critères :

- **Quelque chose que l'on est :** C'est le champ de la biométrie, comprenant des techniques comme la prise d'empreintes digitales, l'analyse rétinienne, l'analyse de la voix, la forme du visage, etc.....
- **Quelque chose que l'on sait :** C'est le système de mot de passe traditionnel.
- **Quelque chose que l'on a :** Cela comprend des mécanismes comme des listes de questions-réponses, des one-time pads (blocs à usage unique), des cartes à puces ou mode standard, etc.

I.2.5.2 Sensibilisation du personnel

Les politiques de sécurité informatique des entreprises s'appuient généralement sur des techniques de protection et des plans d'urgence mais négligent souvent un aspect : le personnel. La stratégie idéale dans le domaine de la sécurité informatique ne se limite pas à des techniques de protection et à des consignes complexes. Elle nécessite également une formation appropriée du personnel. Faute d'une sensibilisation de ce dernier, les mesures de sécurité informatique ne sont qu'à moitiés efficaces. [4]

I.2.5.3 Firewall (pare-feu)

Un firewall est outil informatique (matériel et/ou logiciel) conçu pour protéger les données d'un réseau (protection d'un ordinateur personnel relié à Internet par exemple, ou protection d'un réseau d'entreprise). Il permet d'assurer la sécurité des informations d'un réseau en filtrant les entrées et en contrôlant les sorties selon des règles définies par son administrateur. [2]

I.2.5.4 Antivirus

Il s'agit d'un logiciel capable de détecter et de détruire les virus contenus sur un disque. Le logiciel a pour charge de surveiller la présence de virus et éventuellement de nettoyer, supprimer ou mettre en quarantaine le ou les fichiers infectés. Ils surveillent tous les espaces dans lesquels un virus peut se loger. [2]

I.2.5.5 VPN

(Virtual Private network) : Est un service qui permet à un ou plusieurs postes distants d'établir des connexions privées sécurisées dans le réseau public comme internet, pour communiquer de manière sure. Ce type de liaison est apparu suite à un besoin croissant des entreprises de relier les déférents sites. [2]

I.2.5.6 IDS

La détection d'intrusion est définie comme étant un mécanisme écoutant le trafic réseau de manière furtive, afin de repérer des activités anormales ou suspectes et permettant ainsi d'avoir une stratégie de prévention sur les risques d'attaques. Il existe différents types d'IDS, que l'on classe comme suit : [4]

- ✓ **Système de détection d'intrusion réseau (NIDS) :** un NIDS analyse de manière passive les flux transitant sur le réseau et détecte les intrusions en temps réel, en d'autres termes, un NIDS écoute tout le trafic réseau, puis analyse et génère des alertes si des paquets semblent dangereux. [4]
- ✓ **Système de détection d'intrusion de type hôte (HIDS) :** un HIDS est généralement placé sur des machines sensibles, susceptibles de subir des attaques et possédant des données sensibles pour l'entreprise. [4]
- ✓ **Système de détection d'intrusion de type hybride :** il s'agit d'un système capable de réunir des informations provenant d'un système HIDS ainsi que d'un NIDS. Généralement utilisé dans un environnement décentralisé, il permet de réunir les informations de diverses sondes placées sur le réseau. [4]

I.2.5.7 IPS [4]

L'IPS est un Système de Prévention/Protection contre les intrusions et non plus seulement de reconnaissance et de signalisation des intrusions comme la plupart des IDS. La principale différence entre un IDS (réseau) et un IPS (réseau) tient principalement en deux caractéristiques :

- Le positionnement en coupure sur le réseau de l'IPS et non plus seulement en écoute sur le réseau pour l'IDS (traditionnellement positionné comme un sniffer sur le réseau).
- La possibilité de bloquer immédiatement les intrusions et quel que soit le type de protocole de transport utilisé et sans reconfiguration d'un équipement tierce, ce qui induit que l'IPS est constitué en natif d'une technique de filtrage de paquets et de moyens de blocages.

I.2.5.8 Serveur proxy

Un serveur proxy appelé aussi serveur mandataire, est un composant logiciel informatique qui joue le rôle de l'intermédiaire entre deux machines pour surveiller leurs échanges. La plupart du temps le serveur proxy est utilisé pour le web, il s'agit alors d'un proxy HTTP. Toutefois il peut exister des serveurs proxy pour chaque protocole applicatif (FTP, ...etc.). [5]

I.2.5.9 Cryptographie

La cryptographie est l'étude des méthodes permettant de transmettre des données de manière confidentielle. Afin de protéger un message, on lui applique une transformation qui le rend incompréhensible, c'est ce qu'on appelle le chiffrement, qui, à partir d'un texte en clair, donne un texte chiffré. Inversement, le déchiffrement est l'action qui permet de reconstruire le

texte en clair à partir du texte chiffré en utilisant une clé particulière et un algorithme de déchiffrement. [3]

I.2.5.10 Contrôle d'accès

L'accès au système d'information exige une identification et une authentification préalable. L'utilisation de comptes partagés ou anonymes est interdite. Des mécanismes permettant de limiter les services, les données, les privilèges auxquels à accès l'utilisateur en fonction de son rôle dans l'organisation doit être mis en œuvre. [7]

I.3 Le contrôle d'accès

I.3.1 Définition de contrôle d'accès

Le contrôle d'accès consiste à vérifier si une entité que ce soit une personne, un ordinateur demandant d'accéder à une ressource a les droits nécessaires pour le faire. Avec le contrôle d'accès, on s'intéresse à garantir deux propriétés fondamentales :la confidentialité et l'intégrité des informations contenues dans un système informatique. [1]

L'une des exigences majeures du partage des données entre plusieurs utilisateurs est la protection de ces données contre des atteintes à la confidentialité (divulgation d'information non autorisées), contre des atteintes à l'intégrité (modification non autorisées) et contre des atteintes à la disponibilité (déni de service). Afin d'assurer cette protection, chaque accès aux données doit être contrôlé et bien évidemment tous les accès non autorisés doivent être impérativement bloqués. Cela est appelé le contrôle d'accès. [1]

I.3.2 Définition d'une politique de contrôle d'accès

Elle définit les règles selon lesquelles le contrôle d'accès doit être régularisé. Les politiques sont une exigence de haut niveau qui précisent d'une part, la façon dont le contrôle d'accès est structuré, et d'autre part, la façon dont le contrôle d'accès est structuré, et d'autre part quel utilisateur peut effectuer telles actions sur telles ressources. [8]

À partir de cette définition nous distinguons trois concepts fondamentaux d'une politique de contrôle d'accès :

- Sujet : c'est une entité active qui représente les utilisateurs dans un système. Un sujet est noté **S**. Le sujet est généralement une personne, une application, un processus, une adresse IP…
- Objet : entité passive qui représente les données à protéger contenues dans le système. Un objet est noté **O**. L'objet peut être un fichier, ressource, table relationnelle programme, information…
- Action : représente l'opération possible appelé par le sujet sur l'objet. Une action est notée **A**. L'action peut être lire, écrire, exécuter, modifier, supprimer…

I.3.3 Les modèles de contrôle d'accès

Un modèle de contrôle d'accès peut être défini comme un formalisme (souvent mathématique) qui permet de développer et spécifier le comportement d'un système de manière exacte et de mieux le comprendre. Il permet aussi d'abstraire, donc de faciliter la compréhension d'une politique de sécurité et d'implémenter des mécanismes pour assurer certains objectifs de sécurité. [9]

Plusieurs modèles ont été proposés pour répondre aux différents besoins de contrôle d'accès des applications :

I.3.3.1 DAC (Discretionary Access Control)

Dans Discretionary Access Control ou « modèle de contrôle d'accès discrétionnaire » chaque objet ou ressource du système a un propriétaire (un sujet), lequel peut déterminer les privilèges d'accès à cet objet. Le sujet a un contrôle complet sur tous les objets qui lui appartiennent, il peut changer les permissions d'accès, transférer des objets authentifiés ou des accès à l'information à d'autres sujets. C'est pourquoi il est dit discrétionnaire. Dans ce modèle, les autorisations sont attribuées directement à des sujets en fonction de leur identité. [8]

L'inconvénient d'une telle approche est que, dans les grands systèmes, déterminer l'octroi de l'autorisation sur une ressource donnée à des utilisateurs individuels, est laborieux et difficile à gérer. La révocation de la permission est également complexe lorsque l'utilisateur quitte l'entreprise ou change de fonction, par exemple. L'information peut être copiée d'un objet à un autre, de sorte que l'accès à une copie est possible même si le propriétaire initial ne donne pas accès à l'originale. Puisque les politiques du DAC peuvent être facilement modifiées par le propriétaire, un programme malveillant s'exécutant en son nom pourra aussi changer ces mêmes politiques, ce qui constitue une faiblesse de ce système. [8]

En résumé : [10]

- Le système = l'ensemble (Objet + Sujet).
- Objet et Sujet sont énumérés.
- Autoriser (s, o) = {Ɵ, Lire, Ecrire, Exécuter…} /s ∈ Sujet, o ∈ Objet.
- Représentation :
  - Une matrice (n, m) tel que n=card (Sujet) ; m=card (Objet).
  - Matrice (i, j) = {Ɵ, Lire, Ecrire, Exécuter…}

La matrice de contrôle d'accès est une structure qui contient une ligne par sujet et une colonne par objet dans le système. L'intersection d'une ligne et d'une colonne décrit les droits d'accès du sujet sur l'objet (Lire, Ecrire, Exécuter, …). La figure suivante montre un exemple de matrice de contrôle d'accès :

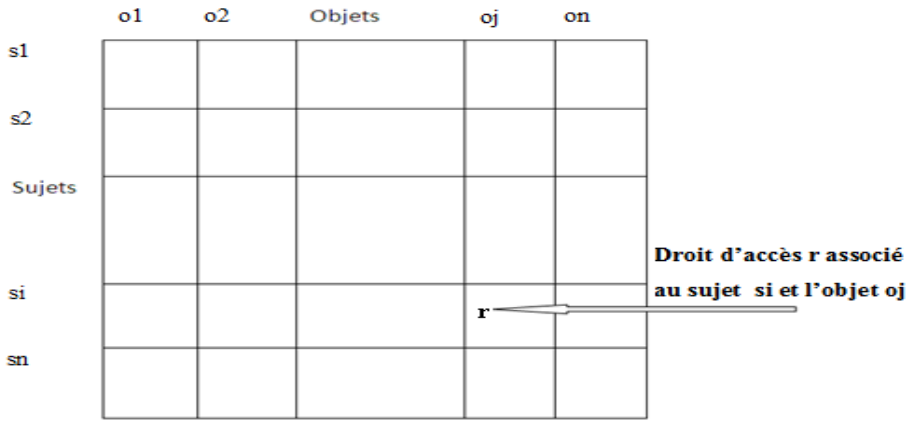

Figure 1 : Matrice de contrôle d'accès [9]

I.3.3.2 MAC (Mandatory Access Control)

Le contrôle d'accès discrétionnaire est généralement défini par opposition au contrôle d'accès obligatoire ou MAC (Mandatory Access Control), qui impose des règles incontournables garantissant l'atteinte des objectifs de sécurité visés. Dans ce type de contrôle d'accès les sujets ne peuvent pas intervenir dans l'attribution des droits d'accès. Ce contrôle d'accès est plus rigide que le contrôle d'accès discrétionnaire mais est, cependant, plus sûr. [11]

Le MAC est utilisé lorsque la politique de sécurité des systèmes d'information impose que les décisions de protection ne doivent pas être prise par le propriétaire des objets concernés, c'est à dire que les décisions de protection doivent lui être imposées par le système. Les sujets et les objets sont classés sur la base des niveaux de sécurité prédéfinis qui sont utilisés dans le procède de décision d'accès. [10]

Quatre niveaux de sécurité sont généralement considérés : Très Secret (TS), Secret (S), Confidentiel (C), et non classifiés (U), avec l'ordre suivant : TS > S > C > U. [11]

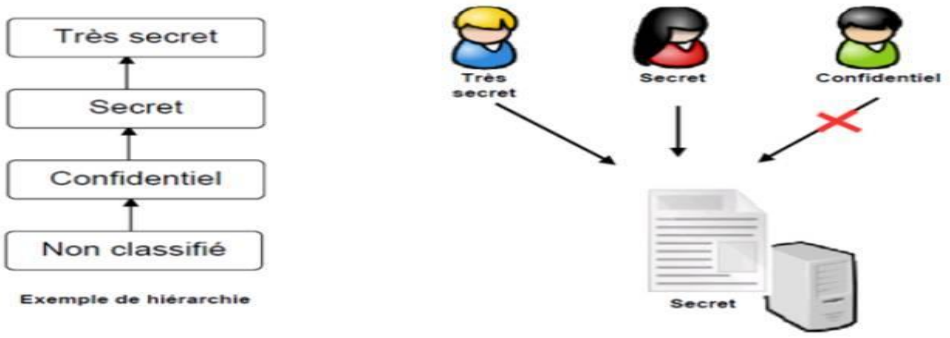

Figure 2 : Exemple de contrôle d'accès avec les niveaux de sécurité [1]

Un sujet ne peut accéder un objet que si sa clearance (autorisation) est supérieure ou égale à la classe de l'objet. [10]

I.3.3.3 RBAC (Role-Based Access Control)

Role Based Access Control (RBAC) ou le contrôle d'accès basé sur les rôles peut être considéré comme une approche alternative au contrôle d'accès obligatoire (MAC) et le contrôle d'accès discrétionnaire (DAC). Dans ce modèle, des permissions qui peuvent être représentées comme un couple (o, a) avec o ∈ O et a ∈ A), sont affectées à des rôles spécifiques au lieu d'être affectés directement à des sujets comme c'est le cas des modèles précédents. Ensuite, les sujets peuvent être attribués aux rôles qui découlent généralement de la structure d'une organisation. Ce modèle simplifie les opérations telles que l'ajout ou la suppression d'un sujet. En effet, les permissions ne sont pas attribuées aux sujets séparément et les sujets ne peuvent acquérir ces permissions qu'à partir de leurs rôles ce qui fait que RBAC est considéré comme un système « idéal » pour les entreprises dont la fréquence de changement du personnel est élevée. En effet quand un gérant Marc est remplacé par la gérante Marie, il n'est pas nécessaire d'affecter à Marie individuellement toutes les permissions de Marc mais il suffit d'affecter à Marie le même rôle de Marc. [9]

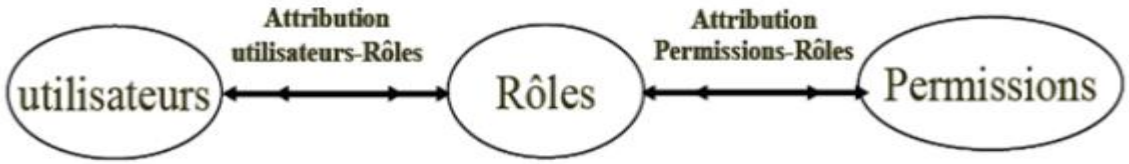

Figure 3 : Attribution des permissions en RBAC [1]

Un rôle peut avoir plusieurs permissions et une permission peut être associée à plusieurs rôles. Un utilisateur peut jouer plusieurs rôles et un rôle peut être attribué à plusieurs utilisateurs [1]. Un sujet a une permission **P** si et seulement si ce sujet est attribué à un rôle qui détient cette permission **P**. [9]

I.3.3.4 Modèle de contrôle d'accès à basé d'équipe (TMAC)

Le modèle de contrôle d'accès à base d'équipe (TMAC : Team-based- Access-Control) a été initialement proposé par K.Thomas. L'objectif est de fournir un contrôle d'accès dans le cas d'un travail collaboratif tout en exploitant la flexibilité du modèle de contrôle d'accès à base de rôle (RBAC). L'entité de base de TMAC, équipe ou "team", est une abstraction qui encapsule un ensemble d'utilisateurs, qui ont des rôles différents et qui collaborent dans le but d'accomplir une tâche commune ou d'atteindre un objectif commun. Les utilisateurs affectés à l'équipe devront bénéficier d'un accès à toutes les ressources de l'équipe. Cependant, les permissions exactes de chaque utilisateur sont déterminées par le rôle qu'il joue et l'activité courante de l'équipe. [1]

I.3.3.5 ORBAC (Organization Based Access Control)

Plusieurs modèles de contrôle d'accès ont été proposés : DAC, MAC, RBAC ... Aucun de ces modèles n'est entièrement satisfaisant au regard des difficultés rencontrées pour mettre en œuvre, au sein d'une organisation, une politique de sécurité qui prendrait en compte les points suivants : [12]

- Des règles qui spécifient des permissions ou des interdictions contextuelles
- Des règles qui spécifient des obligations ou des recommandations.
- Des règles spécifiques à l'organisation.

Le modèle OrBAC (Organization Based Acces Control) tente de prendre en compte ces différents points.

Le contrôle d'accès basé sur l'organisation reprend les principes de rôles du modèle RBAC en offrant en plus, la possibilité de modifier la politique de sécurité en fonction d'une circonstance concrète, c'est-à-dire qu'il exprime facilement les permissions qui dépendent d'un contexte. En dehors des permissions, il offre la possibilité d'exprimer des obligations, des interdictions et même des recommandations dépendant bien évidemment des contextes. Il est centré sur le concept d'organisation (groupe structuré d'entités actives), et tous ses autres concepts sont définis par rapport à l'organisation. À partir des relations ternaires (habilite, utilise et considère), il définit les relations qui existent entre les entités du niveau concret (sujets, objets, et actions), du niveau abstrait (rôles, vues et activités) et l'entité contexte. [12]

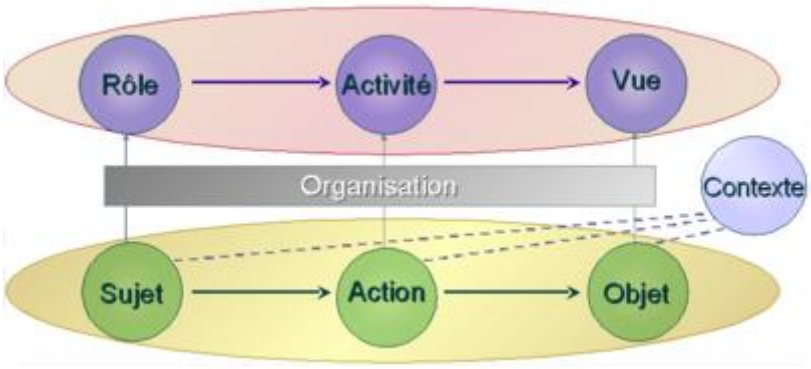

Figure 4 : Modèle ORBAC [13]

OrBAC utilise la notion de hiérarchie de rôle c'est-à-dire un mécanisme d'héritage de permission à travers la hiérarchie de rôle. Ceci peut être applicable dans le cas d'une organisation ayant des sous organisations ; les organisations se succèdent de père en fils. Mais dans le cas des organisations évoluant indépendamment et se situant au même niveau de hiérarchie, on ne peut pas appliquer cette notion de hiérarchie de rôle d'OrBAC mais chercher un autre moyen pour faire une extension. [14]

- **Organisation**

Une organisation peut être définie comme une entité ayant un rôle professionnel ou statutaire bien défini, ou encore, un groupe structuré d'entités actives, c'est-à-dire de sujets (utilisateurs, équipes, ou autres) jouant certains rôles. Il est important de noter qu'un groupe quelconque de sujets n'est pas nécessairement considéré comme une organisation. Autrement dit, le fait que chaque sujet joue un rôle dans l'organisation correspond à un certain accord entre les sujets pour former une organisation. [15]

- **Sujet et rôle**

L'entité sujet est utilisée différemment selon les modèles de sécurité. Dans le modèle OR-BAC, un sujet peut être soit une entité active, c'est-à-dire un utilisateur, soit une organisation. [16]

Les rôles nous permettent de structurer les sujets et de faciliter la mise à jour de la politique de sécurité quand un nouvel utilisateur est ajouté. [16]

Comme les sujets jouent des rôles dans des organisations, nous introduisons une relation entre ces entités : La relation Habilite. Si org est une organisation, S est un sujet et R est un rôle, alors Habilite (org, S, R) signifie que org habilite le sujet S à jouer le rôle R. [16]

Exemple :

- Habilite (org, Mohamed, administrateur) : L'organisation org habilite Mohamed dans le rôle administrateur.

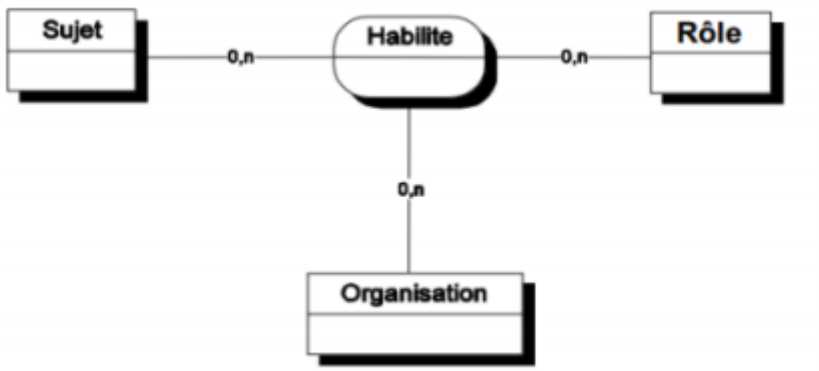

Figure 5 : La relation Habilite [17]

- **Objet et vue**

Le concept de vue permet de modéliser de quelle façon l'organisation utilise les objets, un objet étant une entité non active, comme un fichier, un message électronique, un formulaire imprimé, etc. [18]

Dans la mesure où il est également nécessaire de structurer les objets et d'ajouter de nouveaux objets au système, nous considérons qu'une entité comparable au rôle pour les

sujets est nécessaire pour les objets. Nous l'appelons : entité Vue. De manière intuitive, une vue correspond, comme dans les bases de données relationnelles, à un ensemble d'objets qui satisfait une propriété commune. Par exemple dans un système de fichier administratif, la vue " dossiers administratifs" correspond à l'ensemble des dossiers administratifs des patients, alors que la vue " dossiers médicaux" correspond aux dossiers médicaux des patients. Dans la mesure où les vues caractérisent la manière dont les objets sont utilisés dans l'organisation, nous avons besoin d'une relation qui lie ces trois concepts : la relation Utilise. Si org est une organisation, o est un objet et v est une vue, alors Utilise (org, o, v) signifie que org utilise l'objet o dans la vue v. [12]

Exemple :

- Utilise (org, Fichier.text, dossier administratif) : L'organisation org utilise l'objet Fichier.text comme un dossier administratif.

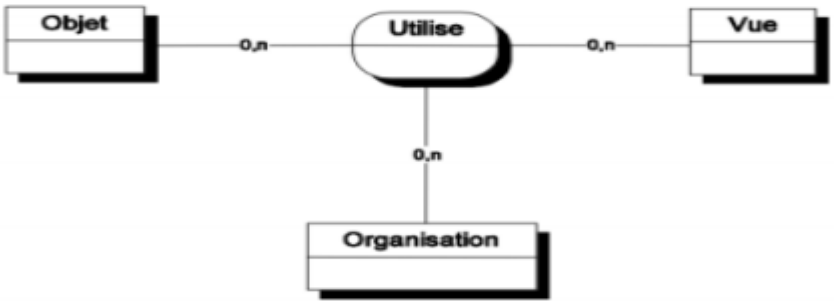

Figure 6 : La relation Utilise [17]

- **Action et activité**

L'entité **Action** englobe principalement les actions informatiques comme "lire", "écrire", "envoyer", etc. L'entité **Activité** correspond à des actions qui ont un objectif commun (exemple : consulter, modifier, transmettre, etc.). [19]

Dans la mesure où des organisations différentes peuvent considérer qu'une même action est employée à la réalisation d'activités différentes, la relation Considère sera utilisée pour associer les entités Organisation, Action et Activité. Plus précisément, si org est une organisation, α est une action et a est une activité, alors Considère (org, α, a) signifie que l'organisation org considère l'action α comme faisant partie de l'activité a. [17]

Exemple :

- Considère (org2, lire, consultation) : « l'organisation org1 considère l'action 'lire' un fichier comme une activité de consultation ».
- Considère (org2, select, consultation) : « l'organisation org2 considère l'action 'select' sur une base de données comme une activité de consultation ».

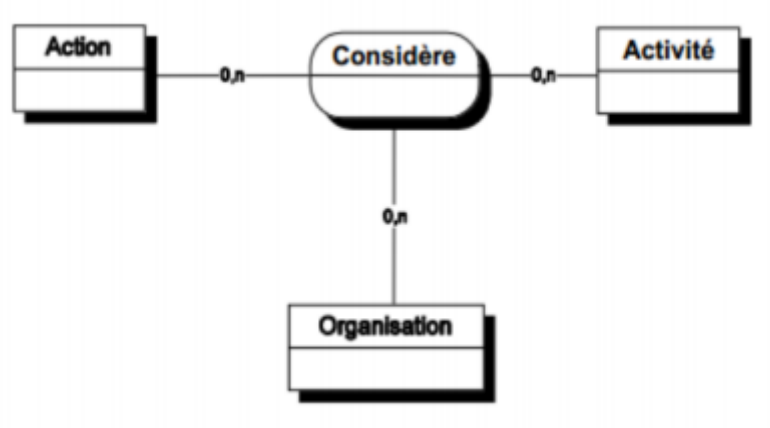

Figure 7 : La relation Considère [17]

- **Contexte**

Les modèles de contrôle d'accès classiques ne permettent de modéliser que des politiques de sécurité qui se restreignent à des permissions statiques. Ils n'offrent pas la possibilité d'exprimer des règles contextuelles. En effet, il est fréquent d'avoir des règles de sécurité spécifiques à un certain contexte. [16]

Les contextes sont utilisés pour spécifier les circonstances concrètes dans lesquelles les organisations accordent aux sujets des permissions de réaliser des actions sur les objets telles que « urgence ». [16]

Les contextes peuvent être vus comme des relations entre les sujets, les objets et les actions définis dans une certaine organisation. Par conséquent, ces quatre entités sont liées par une nouvelle relation appelée Définit. Telle que : *Définit (org, S, α, O, C)* signifie qu'au sein de l'organisation *org*, le contexte *C* est vrai entre le sujet S, l'objet *O* et l'action *α*. [16]

Les contextes peuvent être vus comme des relations ternaires entre les sujets, les objets et les actions définis dans une certaine organisation. Par conséquent, les entités Organisation, Sujet, Objet, Action et Contexte sont liées par une nouvelle relation appelée Définit, telle que : si org est une organisation, s est un sujet, α est une action, o est un objet et c est un contexte, alors Définit (org, s, α, o, c) signifie qu'au sein de l'organisation org, le contexte c est vraie entre le sujet s, l'objet o et l'action α. [17]

Exemple :

- Définit (hôpital général, Marie, lire, carnet.doc, médecin-traitant) qui signifie que l'hôpital général définit Marie ayant le droit de lire carnet.doc car elle est le médecin traitant du patient auquel ce carnet appartient. [19]

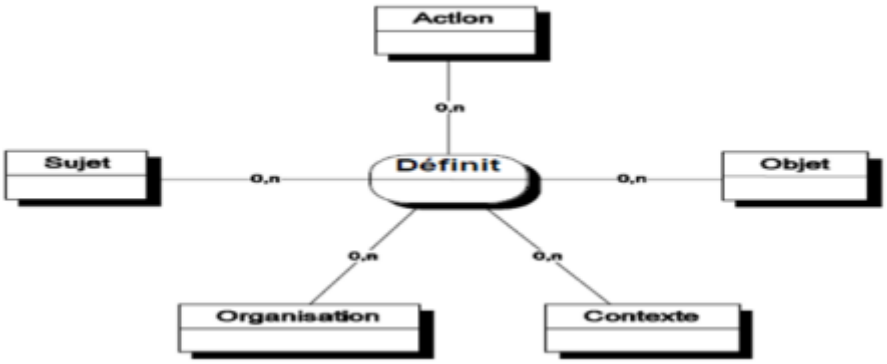

Figure 8 : La relation Définit [17]

- **Expression de politiques de sécurité dans le modèle OrBAC**

Une autre caractéristique d'OrBAC est que les règles exprimées dans ce modèle peuvent définir des permissions, des interdictions, des obligations et des recommandations. Ce modèle est donc beaucoup plus puissant qu'un simple modèle de contrôle d'accès. Ces règles sont de la forme Permission| Interdiction| Obligation |Recommandation (org ; r ; v ; a ; c), où org est une organisation, *r un rôle, v une vue, a une activité et c un contexte* [18]. Par exemple Permission (org ; r ; v ; a ; c) indique que dans une organisation Org, un rôle r est autorisé à effectuer une activité a sur une vue v dans un contexte c.

Les autorisations concrètes de type (sujet, action, objet) quant à eux sont dérivées des relations Est_permis, Est_interdit, Est_obligatoire et Est_recommandé. On aura par exemple Est_permis (s, a, o) qui signifie que le sujet s a la permission de réaliser l'action a sur l'objet o. La figure ci-dessous résume donc le modèle de sécurité OrBAC. [19]

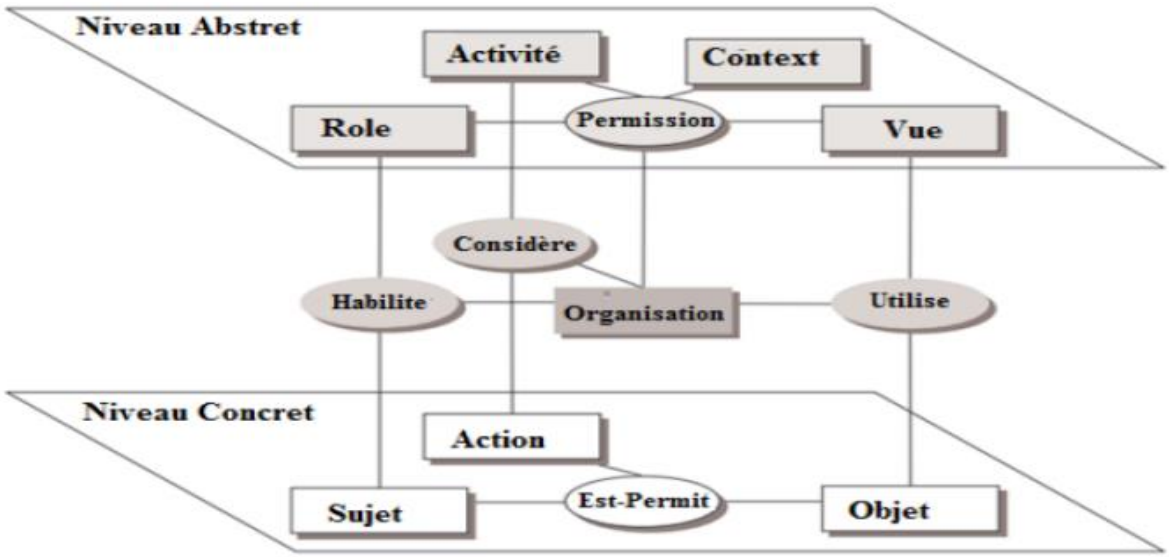

Figure 9 : Structure du modèle OrBAC [12]

I.3.4 Comparaison entre les différents modèles de contrôle d'accès

Le Tableau suivant présente une comparaison entre le modèle de contrôle d'accès ORBAC et les autres modèles de contrôle d'accès (RBAC, TMAC, DAC et MAC …) par rapport aux critères suivants : flexibilité, complexité, support de la collaboration, granularité et utilisation des informations contextuelles.

Critère	**RBAC** [20]	**TMAC** [20]	**ORBAC**	**DAC**	**MAC**
Flexibilité	Oui	Oui	Oui	Oui	Non
Complexité [20]	Moyen	Moyen	Moyen	Moyen	Elevé
Support de la collaboration (groupes d'utilisateurs) [20]	Oui	Oui	Oui	Oui	Oui
Granularité [20]	Faible	Moyen	Elevé	Faible	Moyen
Utilisation des informations contextuelles [20]	Faible	Moyen	Elevé	Non	Non

Tableau 1 : Comparaison entre les modèles de contrôle d'accès

I.4 Conclusion

L'information dans l'entreprise a une importance capitale, ce qui rend sa protection une fonction nécessaire. Il faut noter que la sécurité ne peut être assurée à cent pour cent, mais avec une bonne politique de sécurité, elle peut être très proche à des niveaux acceptés. L'un des aspects de la sécurité informatique est le contrôle d'accès, qui consiste à accorder des droits d'accès selon des contraintes et des règles déjà vérifiées.

Dans ce chapitre, nous avons présenté des généralités sur la sécurité informatique, ses objectifs ainsi que quelques attaques et techniques de défense. Ensuite nous avons parlé de contrôle d'accès, on a présenté quelques modèles de contrôle d'accès de base, puis on a fait une comparaison entre ces différents modèles par apport à quelques critères donnés.

Chapitre II : L’apprentissage automatique

II.1 Introduction

L'intelligence artificielle et la machine learning sont de plus en plus utilisés dans le domaine de la sécurité informatique. L'apprentissage automatique vise à mettre en œuvre des systèmes intelligent en développant des méthodes permettant à un ordinateur de traiter d'énormes quantités de données, d'analyser et de prendre des décisions plus rapidement qu'un humain ne pourrait le faire.

Au cours de ce chapitre nous montrons l'apprentissage automatique et ses principaux concepts.

II.2 Définition de l'intelligence artificielle (IA)

Si nous voulons définir l'intelligence artificielle, nous dirons qu'il n'y a pas un consensus, mais plusieurs définitions existent, selon différentes visions. La définition la plus répandue :

C'est une discipline scientifique relative au traitement des connaissances et au raisonnement dans le but de permettre à une machine d'exécuter des fonctions normalement associées à l'être humain. L'intelligence artificielle tente de reproduire les processus cognitifs humains dans le but de réaliser des actions « intelligente ». Elle est comme « la construction des programmes informatique qui s'adonnent à des tâches qui sont pour l'instant accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : [21]

- L'apprentissage perceptuel.
- L'organisation de la mémoire et le raisonnement critique.

L'intelligence artificielle couvre plusieurs domaines d'applications :

- Le traitement automatique du langage naturel.
- Les systèmes experts et système à base de connaissances.
- Résumer un texte ou le traduire automatiquement.
- Les jeux (échecs, dames …).
- Les interfaces intelligentes.
- L'apprentissage automatique.
- La robotique…

II.3 Définition d'un système intelligent [22]

Un système intelligent est par conséquent un système doté d'une intelligence artificielle. Deux parties se distinguent dans un système intelligent :

- Une partie matérielle composée essentiellement d'un matériel électronique (cartes électroniques programmables, ordinateurs …).
- Une partie informatique composée essentiellement de programmes capables de traiter des informations de différents types.

II.4 Définition de l'apprentissage automatique

La machine Learning ou Apprentissage automatique est un champ de l'intelligence artificielle. Il permet à la machine grâce à l'utilisation massive de données et d'algorithmes d'apprentissage, d'analyser, résoudre des problèmes par elle-même et mettre en œuvre les solutions. [23]

Le principe de base de l'apprentissage automatique est de créer des algorithmes capables de recevoir des données d'entrée et d'utiliser une analyse statistique pour prédire une sortie tout en les mettant à jour à mesure que de nouvelles données deviennent disponibles. [24]

Pour rendre une machine intelligente, il ne suffit pas d'accumuler une masse de connaissances dans celle-ci, mais c'est aussi la doter de capacités d'apprentissage à partir des évènements observés. C'est aussi le pouvoir de la machine de profiter de son expérience, afin d'être mieux préparée à réagir, à l'avenir à des événements similaires. [25]

Toutes les approches d'apprentissage automatique comportent deux phases : la première est celle de l'apprentissage à proprement parler et consiste à choisir un modèle (ex : réseau de neurones) puis ajuster ses paramètres à partir de données en entrée — par exemple des photos de chat et de chien, pour un modèle de reconnaissance visuelle. La deuxième phase est celle de l'inférence. À partir des paramètres qui ont été appris, l'algorithme effectue la tâche qui lui a été fixée — par exemple distinguer les photos de chat des photos de chien. [22]

II.5 Les principales tâches de l'apprentissage automatique

Plusieurs tâches peuvent être associées à l'apprentissage automatique, parmi elles nous citons :

- **La classification**

La classification est le processus de recherche d'un modèle (ou d'une fonction) qui décrit et distingue des classes de données ou des concepts. Le modèle est dérivé sur la base de l'analyse d'un ensemble de données d'apprentissage (c'est-à-dire des objets de données pour lesquels les étiquettes de classe sont connues). Le modèle est utilisé pour prédire la classe d'objets dont l'étiquette de classe est inconnue. [26]

- **L'estimation**

L'estimation consistera à compléter une valeur manquante dans un champ particulier en fonction des autres champs de l'enregistrement [27]. Elle est similaire à la classification sauf que la sortie est une variable numérique plutôt que catégorique.

- **La prédiction**

La prédiction est similaire à la classification et à l'estimation, sauf que pour la prédiction, les résultats sont à l'avenir. Donc toutes les méthodes et techniques utilisées pour la classification et l'estimation peuvent également être utilisées, dans des circonstances

appropriées, pour la prédiction [28]. En effet, la prédiction consiste à prédire la valeur future d'un attribut en fonction d'autres attributs.

- **L'association**

L'association consiste à découvrir des relations intéressantes entre des variables dans des grandes bases de données. Par exemple, les personnes qui achètent une nouvelle maison ont aussi tendance à acheter de nouveaux meubles. Il découvre la probabilité de co-occurrence d'éléments dans une collection. [24]

- **Le clustering**

Le clustering est défini comme une technique d'arrangement d'objets similaires dans une classe appelée cluster. Dans le clustering pas de classification, ni estimation ou prédiction de la valeur des variables cibles mais la segmentation de l'ensemble des données en sous-groupes homogènes. [29]

- **La régression**

La régression, contrairement à la classification, est un processus pour modéliser des fonctions à valeurs continues. Il est utilisé pour prédire les valeurs de données numériques manquantes ou non disponibles plutôt que les étiquettes de classe (discrètes). [30]

II.6 Les types d'apprentissage

Différents algorithmes et techniques tels que la classification, la régression, l'association, les arbres de décision, les machines à vecteurs de support, les réseaux de neurones…etc, sont utilisés pour la découverte de connaissances à partir de bases de données. Cependant, tous les types d'algorithmes peuvent être classés en deux principaux catégories : apprentissage supervisé et apprentissage non supervisé.

II.6.1 L'apprentissage supervisé

L'apprentissage supervisé consiste en des variables d'entrée (x) et une variable de sortie (Y), utiliser un algorithme pour apprendre la fonction de mappage de l'entrée à la sortie **Y=f(X).** Le but est d'appréhender si bien la fonction de mappage que lorsque vous avez de nouvelles données d'entrée (x), vous pouvez prévoir les variables de sortie (Y) pour ces données. [31]

En particulier, l'apprentissage supervisé vise à la modélisation d'une relation entrées-sorties à partir uniquement d'observation de paires entrées-sortie issues de cette relation [21]. Une représentation un peu plus mathématique pour éclaircir le concept : on reçoit des données d'exemple annotées : (x1, y1), (x2, y2), (x3, y3), ... et on espère **prédire** la sortie sur de nouvelles observations : $x* \rightarrow y*$. [32]

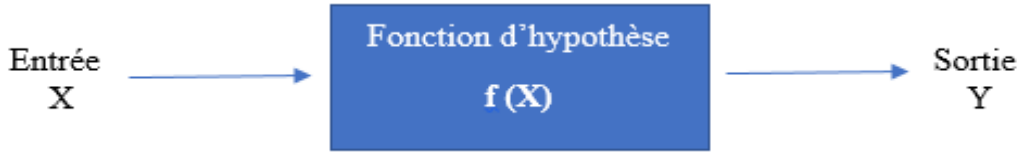

Figure 10 : Apprentissage supervisé

C'est ce qu'on appelle l'apprentissage supervisé, car le processus d'un algorithme tiré de l'ensemble de données, et peut être considéré comme un enseignant supervisant le processus d'apprentissage. Nous connaissons les réponses correctes, l'algorithme effectue des prédictions itératives sur les données d'apprentissage et est corrigé par l'enseignant. L'apprentissage s'arrête lorsque l'algorithme atteint un niveau de performance acceptable [31]. L'apprentissage supervisé englobe les modèles de régression et de classification.

Voici quelques exemples d'algorithmes d'apprentissage automatique supervisé les plus populaires dans la littérature :

- Arbre de décision
- K plus proches voisins (KNN)
- Machine à vecteurs de support (SVM)

II.6.2 L'apprentissage non supervisé

L'apprentissage non supervisé (Unsupervised Learning) consiste à ne disposer que de données d'entrée (X) et pas de variables de sortie correspondantes. [24]

Une représentation un peu plus mathématique pour éclaircir le concept :

Reçoit uniquement des observations brutes de variables aléatoires : x1, x2, x3, x4, ... et on espère découvrir la relation avec des variables latentes structurelles : xi →yi. [32]

L'objectif de l'apprentissage non supervisé est de modéliser la structure ou la distribution sous-jacente dans les données afin d'en apprendre davantage sur les données. On l'appelle apprentissage non supervisé car, contrairement à l'apprentissage supervisé, il n'y a pas de réponse correcte ni d'enseignant. Les algorithmes sont laissés à leurs propres mécanismes pour découvrir et présenter la structure intéressante des données [24]. Les problèmes d'apprentissage non supervisés peuvent être regroupés en problèmes de clustering et d'association. Voici quelques exemples d'algorithmes d'apprentissage automatique non supervisés :

- K-means (K-moyen)
- Réduction de la dimensionnalité
- Clustering hiérarchique

II.7 Les algorithmes de classification

Le data mining offre une très grande variété de techniques et d'algorithmes de fouille de données. Ces algorithmes ont des origines diverses, Certains sont issus des statistiques (régression…), d'autres de l'intelligence artificielle (réseaux de neurones, arbres de décision…), certains encore s'inspirent de la théorie de l'évolution (algorithmes génétiques…). Cette combinaison de technologies facilite la résolution, la compréhension, la modélisation et l'anticipation des problèmes. [33]

Le data mining est un ensemble de techniques complémentaires dédiées à différentes tâches. Ces techniques sont partagées, principalement, entre la classification automatique (supervisée et non supervisée) et la recherche d'associations. [33]

II.7.1 L'algorithme k-Means (Macqueen, 1967)

L'algorithme K-Means (K-moyennes) est le plus connu dans l'Unsupervised Learning. Il s'agit d'un algorithme de Clustering. Ce dernier va mettre dans des "zones" (**Cluster**), les données qui se ressemblent. Les données se trouvant dans le même cluster sont similaires. L'approche de K-Means consiste à affecter aléatoirement des centres de clusters (appelés **centroids**), et ensuite assigner chaque point de nos données au centroid qui lui est le plus proche [34]. Donc chaque cluster de la partition est défini par ses objets et son centroïde. Ensuite, les centres sont redéfinis à partir des objets qui ont été affectés aux différents clusters. Puis, les objets sont assignés en fonction de leur distance aux nouveaux centres et ainsi de suite. L'algorithme se répète donc jusqu'à ce qu'il y ait convergence. La méthode K-moyen est itérative c'est-à-dire qu'elle converge vers une solution quel que soit son point de départ. [35]

Le résultat est un ensemble de clusters compacts et clairement séparés, sous réserve qu'on ait choisi la bonne valeur K du nombre de clusters [36]. On peut résumer le fonctionnement de l'algorithme K-Means dans les étapes suivantes :

1. On choisit K points distincts c_1, …, c_k au hasard parmi $\{x_1, …, x_n\}$ et on considère que c_1, …, c_K c'est les centres des clusters initiales, avec $\{x_1, …, x_n\}$ = les objets (nos données).

2. On répète jusqu'à « stabilisation » des c_k :

2.1. Assigner chaque x_i au cluster $C_k(i)$ tel que dist (x_i, $c_k(i)$) est minimum pour obtenir un ensemble de K classes.

2.2. Recalculer les centres c_k des clusters.

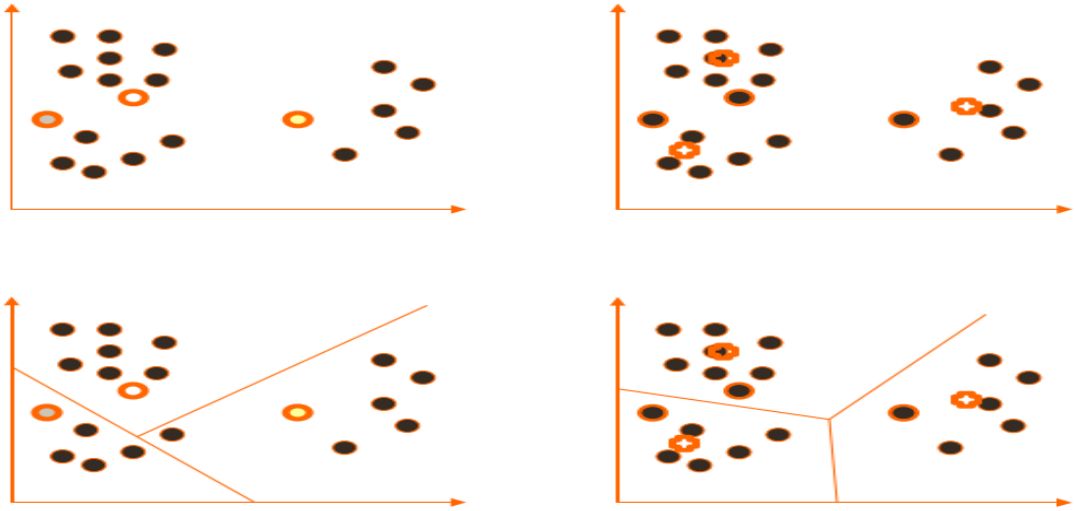

Figure 11 : Exemple sur l'algorithme K-Means [37]

- **La Complexité**

Le nombre de calculs de distance effectués ici est très simple puisque à chaque itération on va calculer la distance de chaque point à chaque centre. [36]

La complexité de l'algorithme du K-moyen est de $O(lkn)$. [35]

Où : l c'est le nombre d'itérations et k c'est le nombre de clusters avec k<n.

- **Les avantages**

  - Simple et facile à implémenter.
  - L'assimilation de cette méthode est rapide. [35]
  - Présente une faible complexité en termes de calcul.
  - Fonctionne avec tous les mesures standards. [38]
  - Insensible à l'ordre des données. [38]

- **Les inconvénients**

  - Nécessite la spécification de nombre de classes à l'avance.
  - Il n'est pas applicable en présence d'attributs qui ne sont pas du type numérique.[38]
  - Le résultat final dépend fortement du choix de centroids initiaux. [38]
  - Ne peut découvrir les groupes non-convexes. [38]
  - Sensible au bruit et aux anomalies.

II.7.2 L'algorithme SVM (Support Vector Machine)

Une machine à vecteurs de support, est un algorithme d'apprentissage automatique supervisé qui peut être utilisé à des fins de classification et de régression. Les SVM sont plus généralement utilisés dans les situations de classification. Les SVM reposent sur l'idée de trouver un hyperplan qui divise au mieux un jeu de données en deux [39] de façon à ce que les données deviennent linéairement séparables.

La technique de construction de l'hyperplan optimal est utilisée pour calculer la fonction de classement séparant les classes tels que : [40]

- ✓ Les vecteurs appartenant aux différentes classes se trouvent de différents côtés de l'hyperplan. [40]
- ✓ La plus petite distance entre les vecteurs et l'hyperplan (la marge) soit maximale. [40]

Donc elles reposent sur deux notions principales : la notion de marge maximale et la notion de fonction noyau. Si on arrive à trouver un séparateur linéaire c'est-à-dire qu'il existe un hyperplan séparateur alors le problème est dit linéairement séparable sinon il n'est pas linéairement séparable et il n'existe pas un hyperplan séparateur. [35]

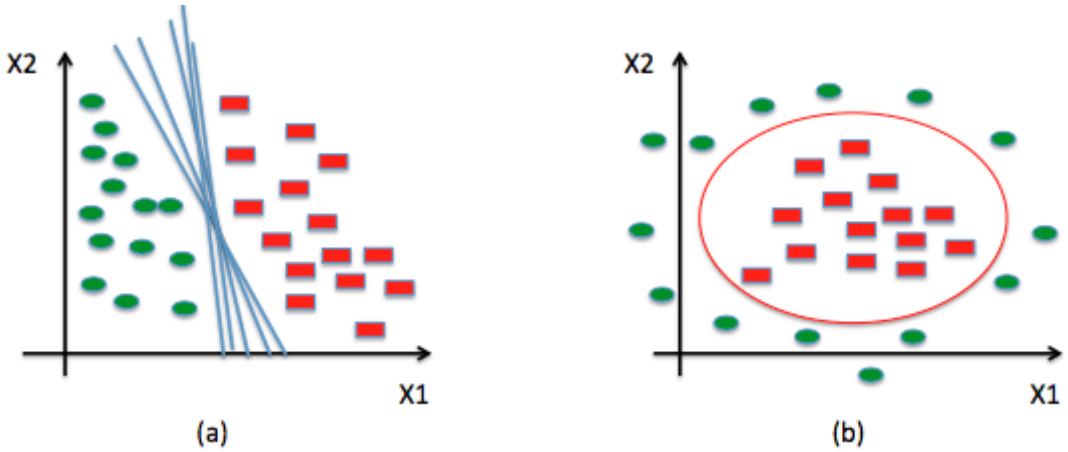

Figure 12 : Problème de classification à deux classes avec une séparatrice linéaire et non linéaire [41]

Dans le cas d'un problème linéairement séparable on peut trouver un ou plusieurs séparateurs linéaires.

S'il y a beaucoup de séparateurs linéaires possibles. Les SVM choisissent seulement celui qui est optimal, c'est-à-dire la recherche d'une surface de décision qui soit éloignée au maximum de tout point de données. Cette distance de la surface de décision au point de données le plus proche détermine la marge maximale du classifieur. En effet, pour obtenir un hyperplan optimal, il faut maximiser la marge entre les données et l'hyperplan. [35]

Pour résoudre le problème de la non linéarité séparatrice, l'idée des SVM est d'augmenter la dimension d'espace de données. Dans ce cas, il est alors probable qu'il existe un séparateur

linéaire. En effet, la chance de trouver un hyperplan séparateur augmente proportionnellement avec la dimension d'espace de données. [35]

- **Avantages**

Parmi les avantages des SVM on cite :

- Elles ont une base théorique solide. [35]
- Elles peuvent être plus efficace car elles utilisent un sous-ensemble de points d'entraînement. [39]
- Les exemples de test sont comparés juste avec les supports vecteur (les points de données les plus proches de l'hyperplan) et non pas avec toutes les données.

- **Inconvénients**

Parmi les inconvénients des SVM on cite :

- Elles utilisent des fonctions mathématiques complexes pour la classification.[35]
- Ne convient pas à des jeux de données plus volumineux, car le temps d'entraînement avec les SVM peut être long. [39]
- Moins efficace sur les jeux de données contenant de bruits et beaucoup d'outliers. [39]

II.7.3 L'algorithme KNN

K Nearest Neighbor (PPV Plus Proches Voisins) est une méthode dédiée à la classification qui peut être étendue à des tâches d'estimation. La méthode PPV est une méthode de raisonnement à partir de cas. Elle part de l'idée de prendre des décisions en recherchant un ou des cas similaires déjà résolus en mémoire.[33]

La décision consiste à chercher les k échantillons les plus voisins de l'objet et de l'affecter à la classe qui est la plus représentative dans ces k échantillons (''dis-moi qui sont tes amis, et je te dirais qui tu es ''). L'approche la plus simple est de rechercher le cas le plus similaire et de prendre la même décision, on parle de 1-NN. Si cette approche peut fournir des résultats acceptables sur des problèmes simples pour lesquels les objets sont bien répartis en groupes denses de même classe, en règle générale, il faut considérer un nombre de voisin plus important pour obtenir de bons résultats. [1]

Contrairement aux autres méthodes de classification il n'y a pas d'étape d'apprentissage consistant en la construction d'un modèle à partir d'un échantillon d'apprentissage. C'est l'échantillon d'apprentissage, associé à une fonction de distance et d'une fonction de choix de la classe en fonction des classes des voisins les plus proches, qui constitue le modèle. [33]

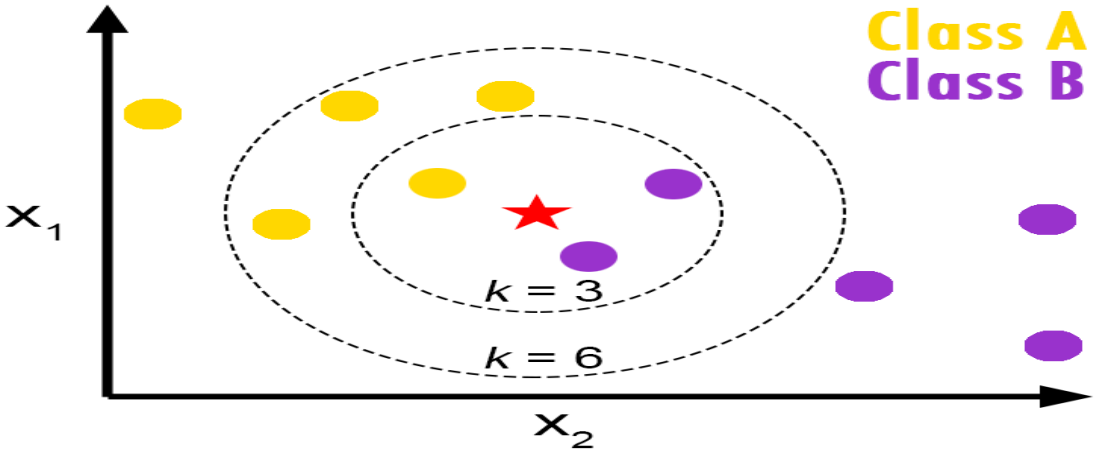

Figure 13 : Exemple d'algorithme KNN [42]

- **Définition de la distance : [35]**

Elle permet de mesurer le degré de différence entre deux vecteurs. Il existe plusieurs types de distance parmi lesquels on trouve :

❖ **La distance Euclidienne :**

$$d(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$$

Où : x, y sont des vecteurs.

❖ **La distance de Minkowsky :**

$$d(x,y)=\left(\sum_{i=1}^{n}|x_i-y_i|^p\right)^{1/p}$$

Où : x, y sont des vecteurs et p : paramètre.

❖ **La distance de Manhattan :**

$$d(x,y)=\sum_{i=1}^{n}|x_i-y_i|$$

Où : x, y sont des vecteurs.

- **L'algorithme des k plus proches voisins** [35]

Pour i = 1 a m faire
Calculer la distance d (Xi, x)
Fin pour
Construire l'ensemble I contenant des indices pour k plus petite distance d (Xi, x)
Retourner Étiquette majoritaire pour {Yi, où $i \in I$}
Où : {$\boldsymbol{X}$: ensemble d'entrainement, $\boldsymbol{Y}$: étiquettes de classe de X, $\boldsymbol{x}$: individu inconnu}

- **Le choix de k**

Le paramètre k doit être déterminé par l'utilisateur : k ∈ N, il est utile de choisir k impair pour éviter les votes égalitaires. Le meilleur choix de k dépend du jeu de données. Il faut tester plusieurs valeurs de k et choisir le k optimal qui donne un meilleur résultat.

- **Avantages de la méthode des k plus proches voisins**

  - Facile à mettre en œuvre et fournit des résultats bons et clairs.
  - La méthode des k plus proches voisins est efficace si les données sont larges et incomplètes. [35]
  - Ne pas faire de l'apprentissage lors de l'introduction de nouveaux attributs. [1]
  - Traiter tout type de données avec un nombre d'attributs élevé. [1]
  - L'algorithme KNN est robuste envers des données bruitées. [35]

- **Inconvénients de la méthode des k plus proches voisins**

  - La faible vitesse de classification due au nombre important de distance à calculer. [43]
  - Cette méthode est gourmande en espace mémoire car elle utilise une grande capacité de stockage pour le traitement des corpus. [35]
  - Nécessite la détermination de k (nombre des plus proches voisins).

II.8 Conclusion

Dans ce chapitre Nous avons parlé de l'apprentissage automatique, ces principaux types ainsi que les algorithmes de classification les plus utilisés.

Dans le chapitre suivant nous allons présenter la modélisation avec la méthode de gestion de risque « EBIOS » et la conception de notre solution.

Chapitre III : Modélisation et Conception

III.1 Introduction

Avant le développement de chaque système d'information il est nécessaire de passer par l'étape de modélisation. L'objectif principal de cette étape c'est maîtriser la complexité d'un système. Dans ce chapitre nous avons basé sur la démarche EBIOS pour étudier et formaliser les objectifs et les exigences de sécurité du système d'information de notre organisme d'accueil « Ministère de la Poste, des Télécommunications, des Technologies et du Numérique (MPTTN) ». Ensuite, nous avons passé à la modélisation de notre solution.

III.2 La méthode EBIOS (Expression des Besoins et Identification des Objectifs de Sécurité)

III.2.1 Présentation de la démarche EBIOS

EBIOS est une méthode développée et maintenue en 1995 par l'ANSSI (Agence Nationale de la Sécurité des Systèmes d'Information) du secrétariat général de la défense nationale française (SGDN). La méthode consiste à formaliser les objectifs et les exigences de sécurité adaptés au contexte du système étudié. [44]

EBIOS a la particularité d'être disponible gratuitement, pour tout organisme souhaitant mener une étude des risques SSI, et mettre en place une politique adéquate de sécurité de l'information, largement utilisée dans le secteur public (l'ensemble des ministères et des organismes sous tutelle), dans le secteur privé (cabinets de conseil, petites et grandes entreprises), en France et à l'étranger par de nombreux organismes en tant qu'utilisateurs ou bénéficiaires d'analyses de risques SSI. [45]

La démarche méthodologique permet d'impliquer l'ensemble des acteurs du système d'information dans la problématique de la sécurité [44]. La démarche de la méthode EBIOS se décompose en cinq modules : (Figure 14)

- Étude du contexte.
- Étude des événements redoutés.
- Étude des scénarios de menace.
- Étude des risques.
- Étude des mesures de sécurité.

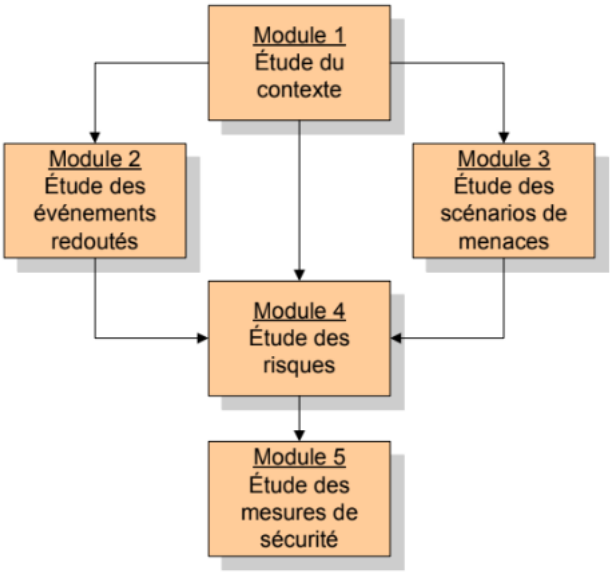

Figure 14 : Démarche EBIOS [44]

III.2.2 Module 1 : Étude du contexte

Cette étape essentielle a pour objectif de comprendre comment réunir les éléments nécessaires pour adapter la gestion des risques au contexte particulier du sujet de l'étude, qui est le système cible [45].

Dans ce module on va décrire le périmètre de l'étude en identifiant les biens essentiels (définis comme l'information ou les processus jugés important pour l'organisme ' immatériels ') ainsi que les biens supports (définis comme des biens sur lesquels reposent les biens essentiels 'physiques').

Au cours de cette étape nous allons passer par les activités suivantes :

- Activité 1 : Étude du système cible

✓ **Présentation de l'organisme MPTTN**

Le Ministère de la Poste et des Technologies de l'Information et de la Communication jusqu'au 25 mai 2017, le Décret exécutif n° 17-272 du 16 Moharram 1439 correspondant au 7 octobre 2017 portant organisation de l'administration centrale du ministère de la poste, des télécommunications, des technologies et du numérique à émané sur la création du MPTTN avec des missions supplémentaires à savoir la numérisation. Le Ministère PTTN se manifeste par sa nature hierachico-fonctionnelle. En effet, par son caractère, elle est organisée hiérarchiquement mais les impératifs de son activité font qu'il est organisé aussi par fonctions et ce, dans un souci de compétitivité et d'efficacité.

✓ **Organigramme de l'entreprise**

Le schéma présenté dans la Figure 2 représente la structure générale de l'entreprise.

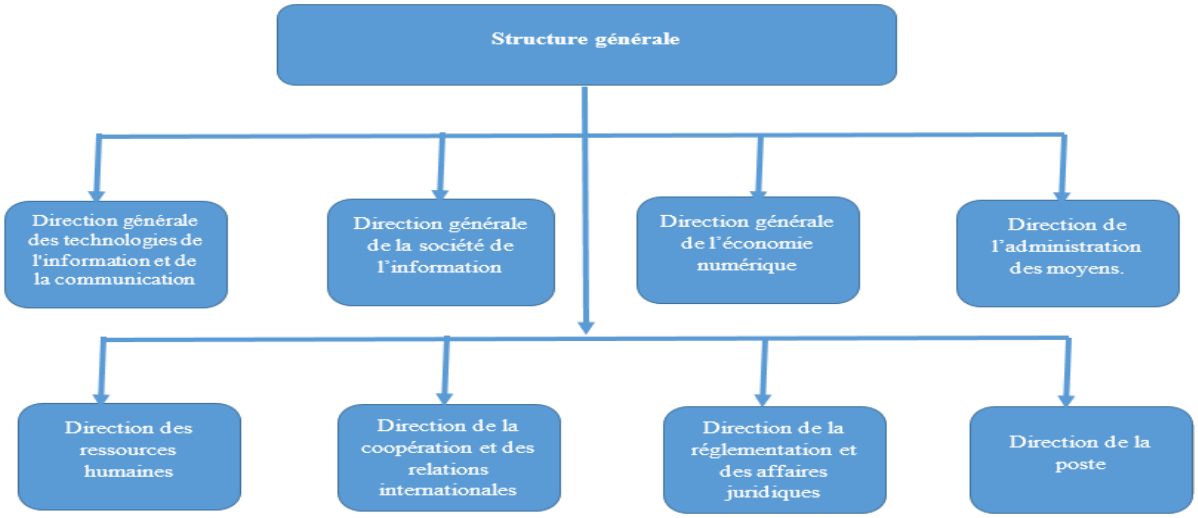

Figure 15 : Organigramme de l'entreprise

La structure générale de l'entreprise **MPTTN** est subdivisée en huit divisions illustrées dans le schéma présenté dans la figure précédente. Chaque une de ces divisions est aussi subdivisée en plusieurs autres sous-divisions.

Nous avons effectué notre stage au sein de la direction de développement et de sécurisation des systèmes d'information, dans **la sous-direction de développement des systèmes d'information et de la numérisation des archives**.

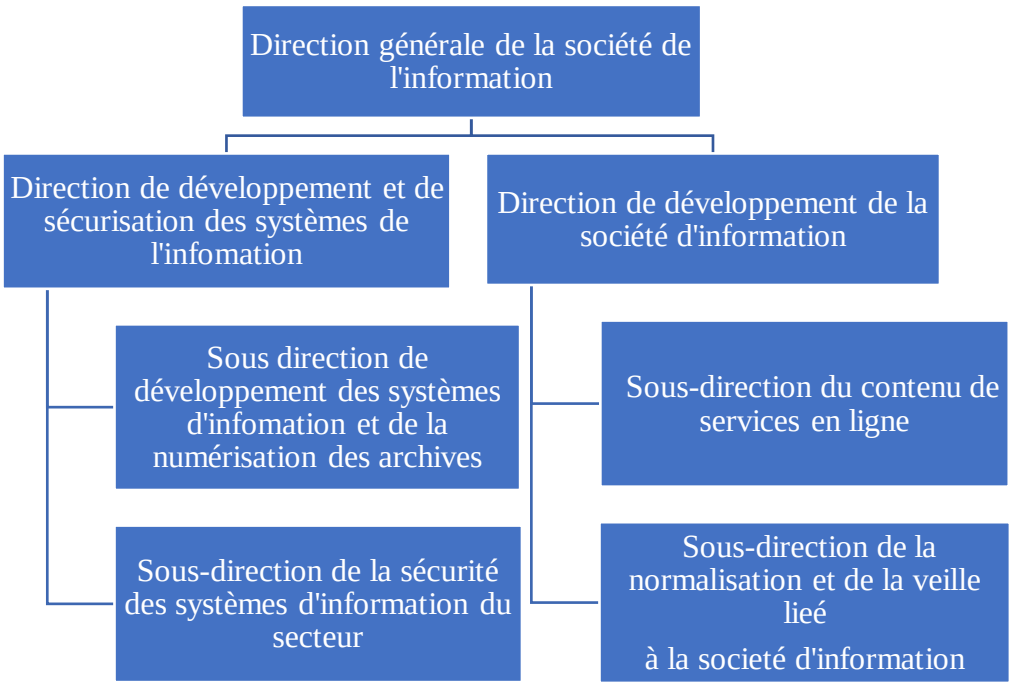

Schéma 1 : Direction générale de la société de l'information

✓ **Présentation de SDDSINA**

La sous-direction de développement des systèmes d'information et de la numérisation des archives « SDDSINA » prend en charge plusieurs missions notamment :

- Identifier les besoins du ministère en matière de logiciels et d'équipements informatiques et de formuler toute proposition au titre de leur mise à niveau.
- Concevoir, de mettre en œuvre et d'administrer les systèmes d'information et les bases de données du ministère.
- Veiller à la maintenance et à la sécurisation des systèmes, des équipements informatiques et des réseaux du ministère.
- Veiller à l'enrichissement du contenu des sites web du ministère et à leur mise à jour.
- Assurer la numérisation et la préservation des archives du ministère et d'entretenir les relations avec le centre des archives nationales.
- Maintenir en condition opérationnelle les équipements informatiques et de communication.

✓ **Description du système informatique**

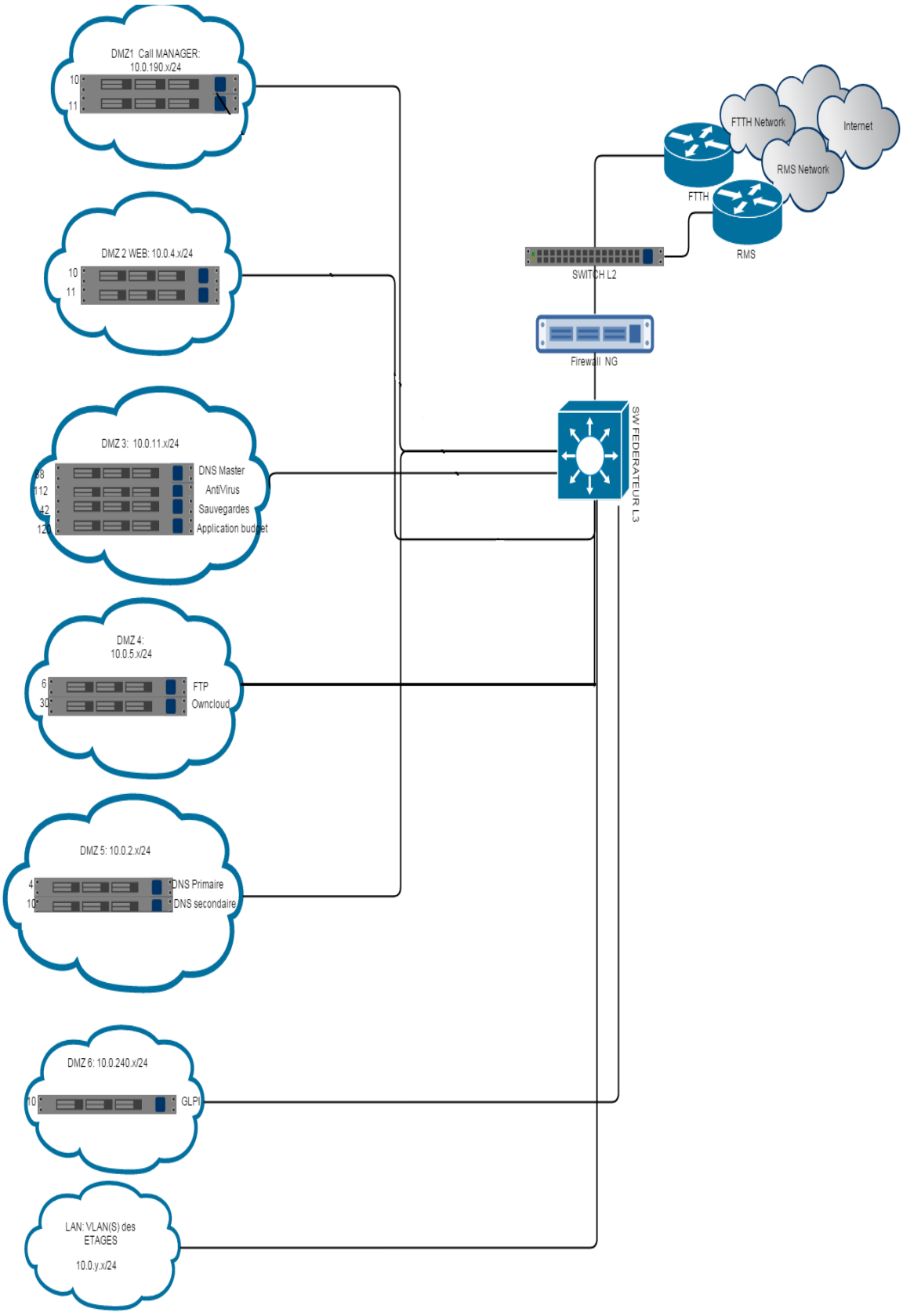

Figure 16 : Architecture du réseau de l'entreprise

L'entreprise MPTTN dispose d'un routeur FTTH qui est connecté directement au réseau FTTH (FTTH Network) et un autre routeur appelé RMS connecté directement au réseau RMS (RMS Network). FTTH Network et RMS Network ce sont deux types de connexion pour exprimer l'internet. Ces deux routeurs sont branchés sur un SWITCH L2 qui est relié avec un firewall NG (Next-generation) ce type de pare-feu fourni une protection adéquate et offre plus de sécurité par rapport au pare-feu traditionnel. Ce dernier est branché sur Switch niveau 3 (SW FEDERATEUR L3) qui permet d'établir un routage et segmenter les réseaux. Chaque service a sa propre réseau LAN et séparé aux autres services par des DMZ.

DMZ (Demilitarized zone) est une zone tampon d'un réseau d'entreprise, située entre le réseau local et Internet, derrière le pare-feu. Il s'agit d'un réseau intermédiaire regroupant des serveurs publics (HTTP, DHCP, mails, DNS, etc.). Ces serveurs devront être accessibles depuis le réseau interne de l'entreprise et, pour certains, depuis les réseaux externes. Le but est ainsi d'éviter toute connexion directe au réseau interne [5]. Pour chaque DMZ on a :

- DMZ1 appelé « Call MANAGER » responsable de gestionnaire d'appel contient deux serveurs avec l'adresse de réseau 10.0.190.x/24.
- DMZ2 appelé « WEB » contient deux serveurs avec l'adresse de réseau 10.0.4.x /24.
- DMZ3 regroupe plusieurs serveurs indépendants (DNS Master(principale), Serveur AntiVirus, Serveur de Sauvegardes, Serveur Application budget) avec l'adresse de réseau 10.0.11.x/24.
- DMZ4 contient deux serveurs avec l'adresse de réseau 10.0.5.x/24 :
  - ✓ Le premier serveur appelé « FTP » : serveur de partage via le protocole FTP (File Transfer Protocol), c'est un protocole de communication destiné au partage des fichiers sur un réseau TCP/IP.
  - ✓ Le deuxième serveur appelé « Owncloud » c'est un serveur de partage collaboratif pour les utilisateurs.
- DMZ5 contient deux serveurs DNS avec l'adresse de réseau 10.0.2.x/24. Le premier appelé « DNS primaire » et le deuxième « DNS secondaire ». Ces serveurs chargent les requêtes au sein de LAN pour optimiser la bande passante, Si le DNS local ne trouve pas la réponse alors, il va contacter DNS Master pour récupérer la bonne information.
- DMZ6 contient serveur « GLPI » (Gestionnaire Libre de Parc Informatique) avec l'adresse de réseau 10.0.240.x/24. Ce système fourni aux utilisateurs un service leur permettant de signaler des incidents ou de créer des demandes basées sur un actif ou non, ceci par la création d'un ticket d'assistance.
- Le dernier nuage représente les VLAN(s) de tous les étages de l'entreprise où chaque étage a un VLAN à part.

- Activité 2 : Sources de menaces retenues

Les sources de menaces retenues peuvent être présentées sous différentes figures, soit de sources humaines ou non humaines. Dans notre cas, nous citons les sources de menaces suivantes [45] :

Hacker (un pirate informatique), Maintenance informatique (mauvaise application de procédure), Employé peu sérieux, Administrateur peu sérieux, Client, Partenaire, Logiciel malveillant (virus, vers, cheval de troie …), Panne électrique, Script-kiddies, Incendie des locaux. Le tableau suivant illustre les types de sources de menaces avec des exemples :

Types de sources de menaces	Retenu ou Non Retenu	Exemples de source de menace
Source humaine interne, malveillante, avec de faibles capacités	Oui	Employé peu sérieux
Source humaine interne, malveillante, avec des capacités illimitées	Oui	Administrateur peu sérieux, Maintenance informatique, partenaire
Source humaine externe, malveillante, avec de faibles capacités	Oui	Script-Kiddies, Client
Source humaine externe, malveillante, avec des capacités illimitées	Oui	Hacker
Source humaine interne, sans intention de nuire, avec de faibles capacités	Oui	Employé peu sérieux
Source humaine interne, sans intention de nuire, avec des capacités importantes	Oui	Administrateur peu sérieux
Source humaine externe, sans intention de nuire, avec de faibles capacités	Oui	Client
Source humaine externe, sans intention de nuire, avec des capacités importantes	Non	
Logiciel malveillant	Oui	Logiciel malveillant (virus, vers, cheval de troie)
Événement interne	Oui	Panne électrique, Incendie des locaux

Tableau 2 : Présentation des sources de menaces

- Activité 3 : les biens supports

Le tableau suivant exprime la liste des biens supports :

Types de biens supports	Biens supports
Matériels	Ordinateurs, routeurs (FTTH, RMS), commutateurs, points d'accès, serveur d'annuaire " Active Directory ", pare-feu, serveurs (serveur FTP, serveur DNS, serveur antivirus...), Câbles.
Logiciels	Systèmes d'exploitation, le SGBD contenant les bases de données, service de messagerie.
Canaux informatiques	L'ensemble des périphériques réseaux utilisés comme Wifi, fibre optique.
Organisations	Fournisseur d'accès Internet, hébergeur, partenaire.
Personnels	Directeurs, sous-directeurs, chefs de service, chefs de divisions, employés, clients.
Supports papiers	Imprimantes, photocopieur, scanner.

Tableau 3 : Présentation des biens support à protéger

- Activité 4 : Les biens essentiels identifiés

- ❖ Gérer le contenu du site web.
- ❖ Gérer les transmissions réseau.
- ❖ Gérer les accès au serveur de courrier.

- Activité 5 : Les sources de risques pertinentes

Les sources de risques sont divisées en trois catégories :

- ❖ Les personnes internes à considérées : Administrateur peu sérieux, maintenance informatique, employé peu sérieux, partenaire.
- ❖ Les personnes externes à considérées : Client, hacker, script-kiddies.
- ❖ Les sources non humaines à considérer : Logiciel malveillant, panne électrique, incendie des locaux.

- Activité 6 : Les métriques utilisées

- ➢ **Les critères de sécurité retenus : [46]**

Afin d'exprimer les besoins de sécurité, les critères de sécurité retenus sont les suivants :

- La disponibilité : propriété d'accessibilité au moment voulu des biens essentiels.
- L'intégrité : propriété d'exactitude et de complétude des biens essentiels.
- La confidentialité : propriété des biens essentiels de n'être accessibles qu'aux utilisateurs autorisés.

- **Échelles à utiliser :**

Les échelles suivantes seront utilisées pour exprimer les besoins de sécurité en termes de disponibilité, d'intégrité et de confidentialité.

Niveaux	Description détaillée de l'échelle
Disponibilité	
Plus de 72h	Le bien essentiel peut être indisponible plus de 72 heures.
Entre 24 et 72h	Le bien essentiel doit être disponible dans les 72 heures.
Entre 4 et 24h	Le bien essentiel doit être disponible dans les 24 heures.
Moins de 4h	Le bien essentiel doit être disponible dans les 4 heures.
Intégrité	
Détectable	Le bien essentiel peut ne pas être intègre si l'altération est identifiée.
Maitrisé	Le bien essentiel peut ne pas être intègre, si l'altération est identifiée et l'intégrité du bien essentiel retrouvée.
Intègre	Le bien essentiel doit être rigoureusement intègre.
Confidentialité	
Public	Le bien essentiel est public.
Limité	Le bien essentiel ne doit être accessible qu'au personnel et aux partenaires.
Réservé	Le bien essentiel ne doit être accessible qu'au personnel (interne) impliquées.
Privé	Le bien essentiel ne doit être accessible qu'à des personnes identifiées et ayant le besoin d'en connaître.

Tableau 4 : Présentation des échelles de disponibilité, d'intégrité et de confidentialité [45]

L'échelle suivante sera utilisée pour estimer la gravité des événements redoutés et des risques : [47]

Niveaux de l'échelle	Description détaillée de l'échelle
1. Négligeable	Surmonter les impacts sans aucune difficulté.
2. Limitée	Surmonter les impacts malgré quelques difficultés.
3. Importante	Surmonter les impacts avec de sérieuses difficultés.
4. Critique	Ne surmonter pas les impacts (sa survie est menacée).

Tableau 5 : Échelle de gravité [47]

L'échelle suivante sera utilisée pour estimer la vraisemblance des scénarios de menaces et des risques : [47]

Niveaux de l'échelle	Description détaillée de l'échelle
1. Minime	Cela ne devrait pas se (re)produire.
2. Significative	Cela pourrait se (re)produire.
3. Forte	Cela devrait se (re)produire un jour ou l'autre.
4. Maximale	Cela va certainement se (re)produire prochainement.

Tableau 6 : Échelle de vraisemblance [47]

- **Activité 7 : les mesures de sécurité existantes**

Il existe plusieurs mesures de sécurité au niveau de l'organisme. Chacune de ces mesures repose sur des biens supports comme il est définit dans le tableau suivant :

Mesures de sécurité existantes	Bien supports lesquelles elles reposent
– Authentification centralisée et sécurisée via un contrôleur de domaine	Serveur d'annuaire "Active Directory"
– Solution Antivirale	Serveur antivirus
– séparation en DMZ – Politique de gestion des ACL	Service de messagerie, pare-feu, routeurs, serveurs (DNS, FTP…)
– Chaque étage a sa propre VLAN	Switch (SW FEDERATEUR L3, Switch de chaque étage)
– Accès sécurisé par mot de passe au serveur de courrier	Serveur de courrier
– Utilisation d'un Pare feu matériel	Par feu (Firewall NG)

Tableau 7: Mesures de sécurité existantes

III.2.3 Module 2 : Étude des événements redoutés

Le deuxième module inclut l'étude des événements redoutés, ces derniers représentant les scénarios génériques qu'il faut éviter sur le périmètre de l'étude. Dans cette méthode, les réflexions sont menées sur les biens essentiels et non pas sur les biens de support. [44]

Chaque ligne du tableau suivant représente un événement redouté qui combine (les sources de menaces, le bien essentiel, le critère de sécurité, le besoin de sécurité concerné, les impacts si le besoin de sécurité n'est pas satisfait et la gravité (échelle de gravité)).

Evènement redouté	Bien essentiel	Critère	Besoin de sécurité	Source de menace	Impact	Gravité
Site internet indisponible	Gérer le contenu du site web	Disponibilité	Entre 4 et 24 h	Logiciel malveillant, Script-kiddies, Panne électrique, Employé peu sérieux	Site web piraté, Perte de crédibilité	Limitée
Modification non autorisée du site internet	Gérer le contenu du site web	Intégrité	Détectable	Script kiddies	Site web piraté, Perte de crédibilité	Limitée
Interception des données transitant sur le réseau	Gérer les transmissions réseau	Confidentialité	Réservé	Logiciel malveillant, Script-kiddies, Hacker	Compromettre la confidentialité des données	Critique

Modification des données transitant sur le réseau	Gérer les transmissions réseau	Intégrité	Détectable	Script-kiddies, Logiciel malveillant, Employé peu sérieux	Compromettre l'intégrité des données	Critique
Usurpation des mots de passe	Gérer les accès au serveur de courrier	Confidentialité	Réservé	Logiciel malveillant, Hacker	Piratage de compte de l'utilisateur et divulgation des données sensibles	Critique

Tableau 8 : Étude des événements redoutés

III.2.4 Module 3 : Étude des scénarios de menaces

L'objectif ici est d'obtenir une liste hiérarchisée de tous les scénarios de menaces possibles. Dans ce module, les scénarios de menaces seront identifiés en termes de vraisemblance (échelle de vraisemblance). Nous complétons le tableau de scénarios de menaces tout en respectant les activités suivantes : [45]

- Déterminer les sources les plus pertinentes en utilisant la liste correspondante. [45]
- Vérifier les vulnérabilités et les compléter si besoin. [45]
- Estimer la vraisemblance des scénarios en utilisant l'échelle définie dans le tableau 6. [45]
- Déterminer les biens supports et le critère de sécurité.

Scénarios de menaces	Bien supports	Critères de sécurité	Sources de menaces	Vraisemblance
Menaces sur la salle de matériels informatiques	Data center	Confidentialité	Employé peu sérieux, Maintenance informatique, Incendie des locaux, Panne électrique	Significative
Menace sur la base de données causant une altération	Base de données	Intégrité	Employé peu sérieux, Administrateur peu sérieux, Logiciel malveillant, Script-kiddies	Forte
Accès aux données sensibles et confidentielles de l'entreprise	Service de messagerie	Confidentialité	Hacker, Script-kiddies, Client	Significative

Interception des communications circulant dans le réseau.	L'ensemble des périphériques réseau utilisés	Confidentialité	Logiciel malveillant, Script-kiddies, Hacker	Significative
Ecoute du trafic réseau en utilisant un sniffeur (analyseur réseau) comme wireshark	L'ensemble des périphériques réseau utilisés	Confidentialité, Intégrité	Hacker, Client ou Employé peu sérieux	Minime
Accès au réseau LAN en passant par un point d'accès non sécurisé	Point d'accès	Confidentialité, Intégrité et Disponibilité	Hacker, Partenaire, Client, Employé peu sérieux	Maximale

Tableau 9 : Étude des scénarios de menaces

III.2.5 Module 4 : Étude des risques

Le quatrième module permet de réaliser l'identification des risques pesant sur le périmètre de l'étude. L'objectif est de créer le lien entre les événements redoutés et les scénarios de menaces, c'est-à-dire entre ce que l'organisme craint et ce à quoi il est exposé. Dans ce même module, on identifie également les objectifs de sécurité afin de choisir la manière dont chaque risque devra être traité en fonction de l'évaluation de son impact et de la potentialité de son occurrence [44]. Le niveau d'un risque est estimé en termes de gravité et de vraisemblance [45] tel que :

- La gravité représente l'ampleur d'un risque. [45]
- La vraisemblance traduit la faisabilité d'un risque. Elle dépend essentiellement des vulnérabilités des supports face aux menaces et des capacités des sources de risques à les exploiter. [45]

D'après l'étude précédente des scénarios de menaces et des évènement redoutés et en tenant compte les mesures de sécurité existantes nous avons peut dégager les risques suivants :

- ✓ Risque lié à l'indisponibilité du site internet.
- ✓ Risque lié à l'altération du contenu du site Internet sans pouvoir la détecter ni la retrouver.
- ✓ Risque lié à l'interception des données circulant dans le réseau.
- ✓ Risque lié à l'usurpation des mots de passe.
- ✓ Risque lié aux pannes dans les salles de matériels informatiques.
- ✓ Accès aux données sensibles et confidentielles de l'entreprise (accès au service de messagerie ou serveur de bases de données).
- ✓ Accès gratuit à internet.
- ✓ Diffusion non autorisée d'informations.

III.2.6 Module 5 : Étude des mesures de sécurité

Ce module consiste à formaliser les mesures de sécurité à mettre en œuvre, afin de déterminer les actions à entreprendre ainsi que les mesures de sécurité adéquates. Des activités d'élaboration et de suivi de la réalisation du plan de traitement sont mises en œuvre à la fin de ce module. [44]

D'après ce que nous avons présenté auparavant et afin de réduire les risques, nous avons proposé les mesures de sécurité suivantes :

- Mettre en œuvre un système de contrôle d'accès pour gérer les accès dans l'entreprise, ce système fonctionne de manière automatique à cause de l'utilisation des techniques de la machine Learning, et un modèle de contrôle d'accès basé sur l'organisation. Il donne l'accès seulement aux personnes autorisées.
- Utilisation des mécanismes cryptographique pour le chiffrement des flux de données circulant dans le réseau et aussi les fichiers et toute information secrète à l'aide de certificats électroniques.
- Mettre en place un système de détection d'intrusion (IDS) pour détecter toute activité suspecte dans le réseau.
- Mettre en place un système de prévention d'intrusion (IPS) qui va réagir en temps réelle en stoppant l'activité suspecte qu'il reconnait notamment.
- Mettre en place un VPN "IPsec ou SSL" qui permet de sécuriser les communications en assurant l'authentification de ses entités et la confidentialité des communications.
- Automatiser l'accès au data center par l'utilisation d'une carte au lieu d'un agent de sécurité.
- Mettre en place un deuxième Pare-feu qui prend la main en cas de panne de premier.
- Mettre en place un SIEM (Security Information and Event Management). Il s'agit d'un outil qui permet d'enregistrer les logs et les analyser afin de surveiller en temps réel les événements informatiques avec un processus préalablement établi.
- Mise en place d'un serveur proxy qui joue le rôle d'intermédiaire en se plaçant entre deux hôtes pour faciliter ou surveiller leurs échanges.
- Mettre en place un serveur Radius permettant l'authentification des différents utilisateurs avant d'accéder à un point d'accès.

III.3 L'approche UML (*Unified Modeling Language*)

III.3.1 Définition

La notation UML est un langage visuel constitué d'un ensemble de schémas, appelés des diagrammes, qui donnent chacun une vision différente du projet à traiter. UML nous fournit donc des diagrammes pour représenter le logiciel à développer : son fonctionnement, sa mise en route, les actions susceptibles d'être effectuées par le logiciel, etc. [48]

III.3.2 Diagramme de cas d'utilisation

Le diagramme de cas d'utilisation est un diagramme UML utilisé pour donner une vision globale du comportement fonctionnel d'un système. Dans notre solution, nous distinguons deux acteurs :

- **Administrateur :** personne qui a pour rôle principal de gérer et administrer toutes les fonctionnalités du système.
- **Utilisateur :** représente l'utilisateur de l'application (Employé).

La figure suivante représente le diagramme de cas d'utilisation de notre système :

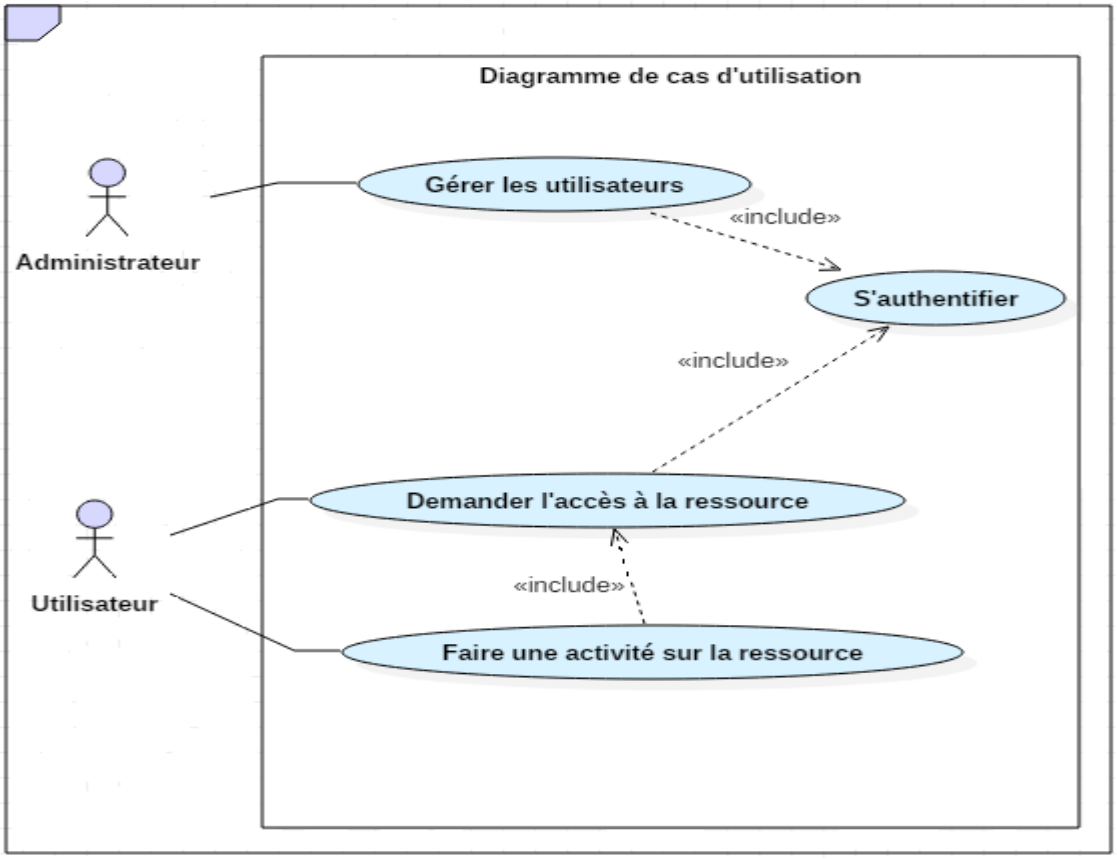

Figure 17 : Diagramme de cas d'utilisation global

Cas d'utilisation	Acteurs	Description
S'authentifier	Administrateur, Utilisateur	L'administrateur, l'utilisateur, doivent s'authentifier pour pouvoir accéder à leur espace.
Gérer les utilisateurs	Administrateur	L'administrateur peut ajouter, modifier, ou supprimer un utilisateur ainsi que désactiver son compte.
Demander l'accès à la ressource	Utilisateur	L'utilisateur fait une demande d'accès à la ressource en remplissant le formulaire d'accès.
Faire une activité sur la ressource	Utilisateur	Dès que l'utilisateur aura l'accès à la ressource, il peut faire les activités suivantes (consultation ou suppression) sur cette ressource.

Tableau 10 : Descriptions des cas d'utilisation du diagramme de cas d'utilisation global

III.3.3 Diagramme de séquence

Le diagramme de séquence représente les interactions entre les objets et les acteurs pour un cas d'utilisation dans un système.

Nous illustrons les diagrammes de séquence suivants : S'authentifier (figure 18), Demander l'accès à la ressource (figure 19) et Faire une activité sur la ressource (figure 20).

III.3.3.1 S'authentifier

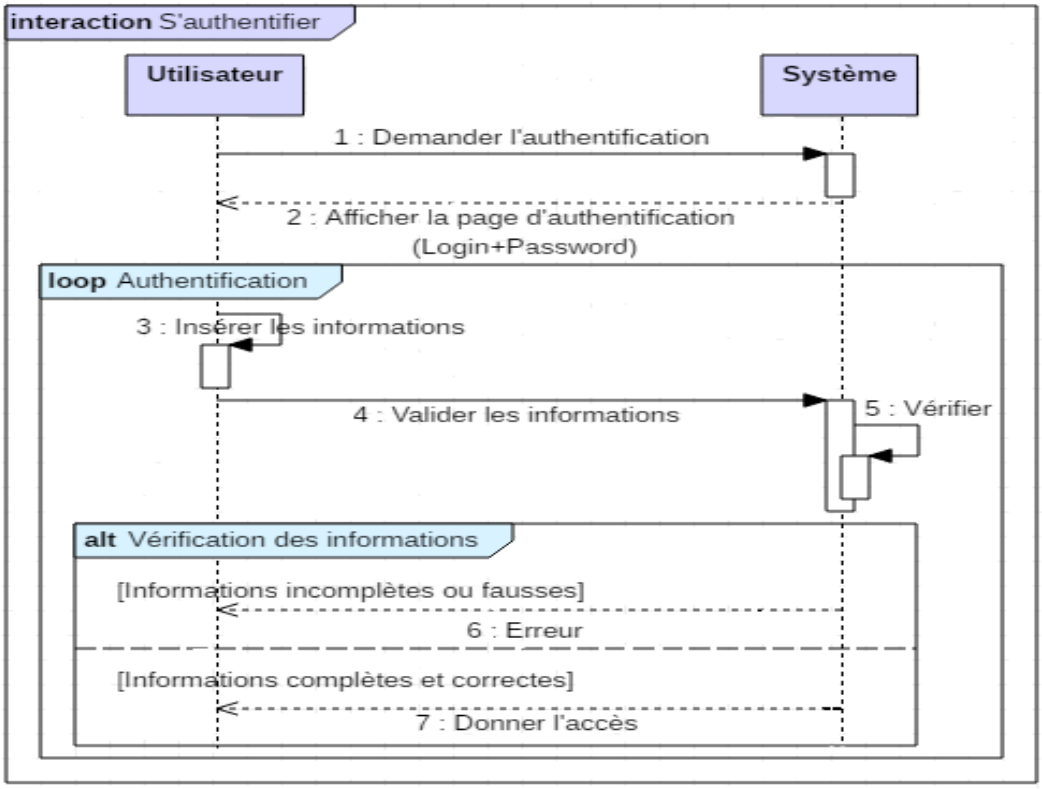

Figure 18 : Diagramme de séquence « S'authentifier »

III.3.3.2 Demander l'accès à la ressource

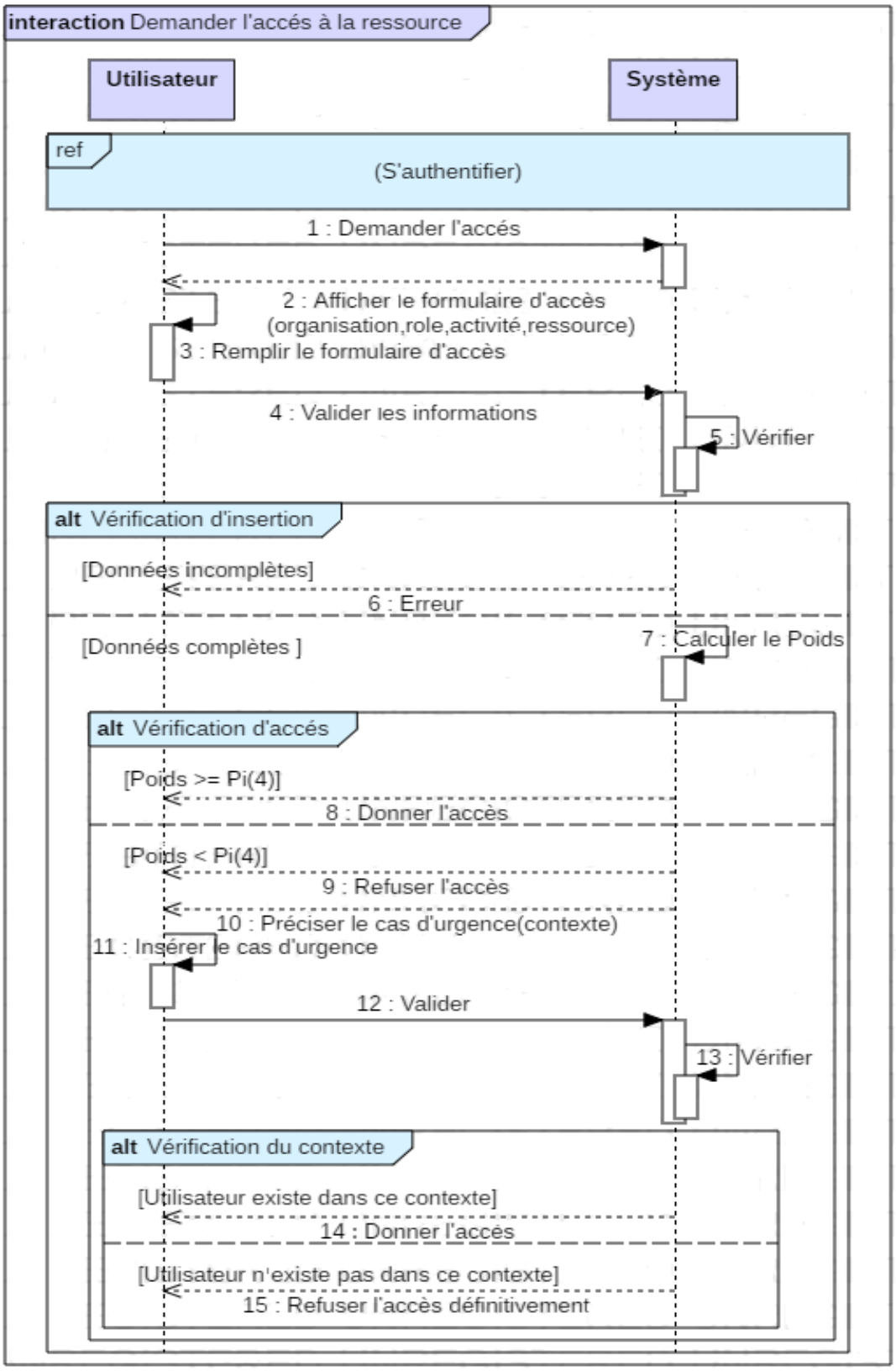

Figure 19 : Diagramme de séquence « Demander l'accès à la ressource »

III.3.3.3 Faire une activité sur la ressource

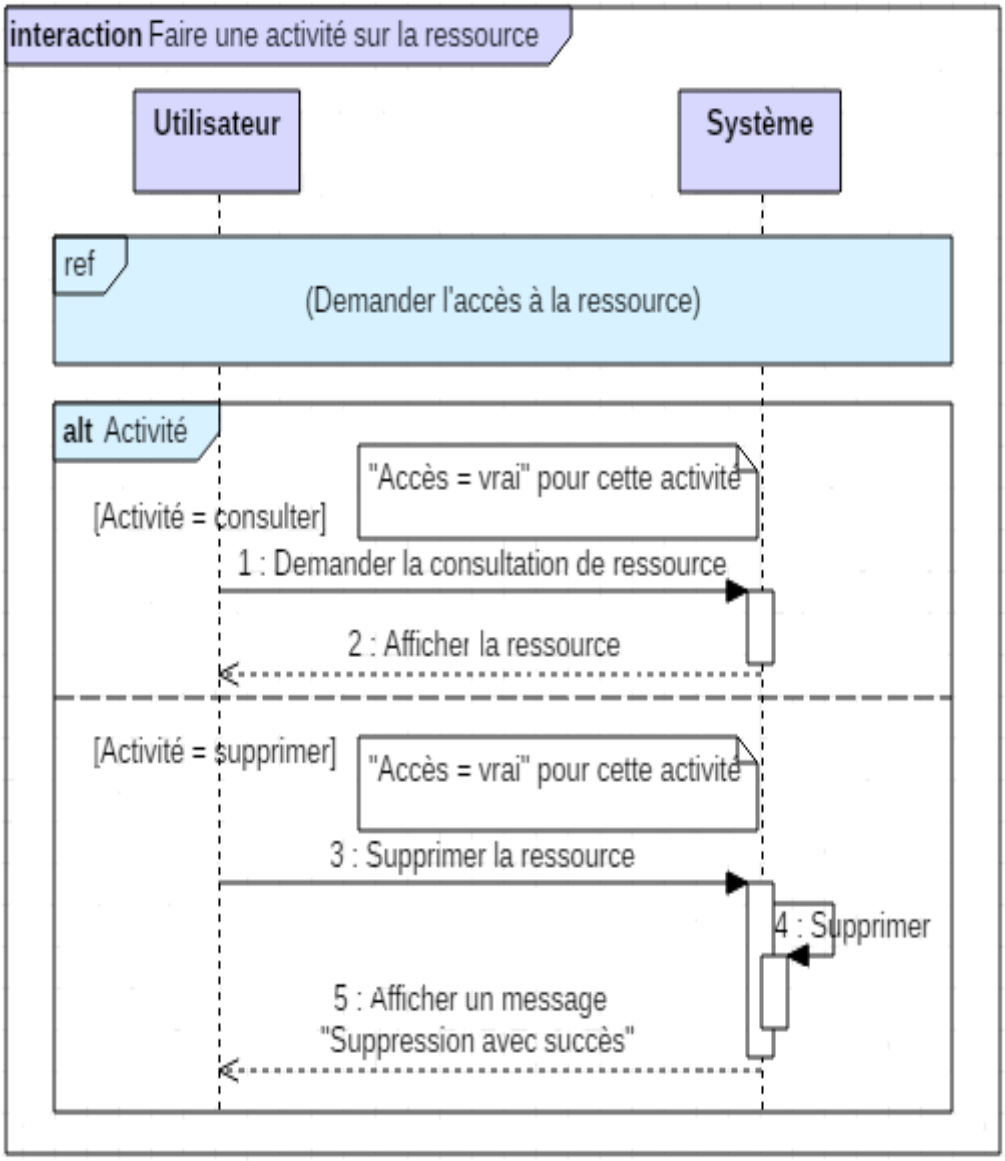

Figure 20 : Diagramme de séquence « Faire une activité sur la ressource »

III.3.4 Diagramme de classe

Le diagramme de classe permet de décrire la structure du système en montrant les classes intervenantes et les relations entre elles. La figure suivante représente le diagramme de classe de notre système :

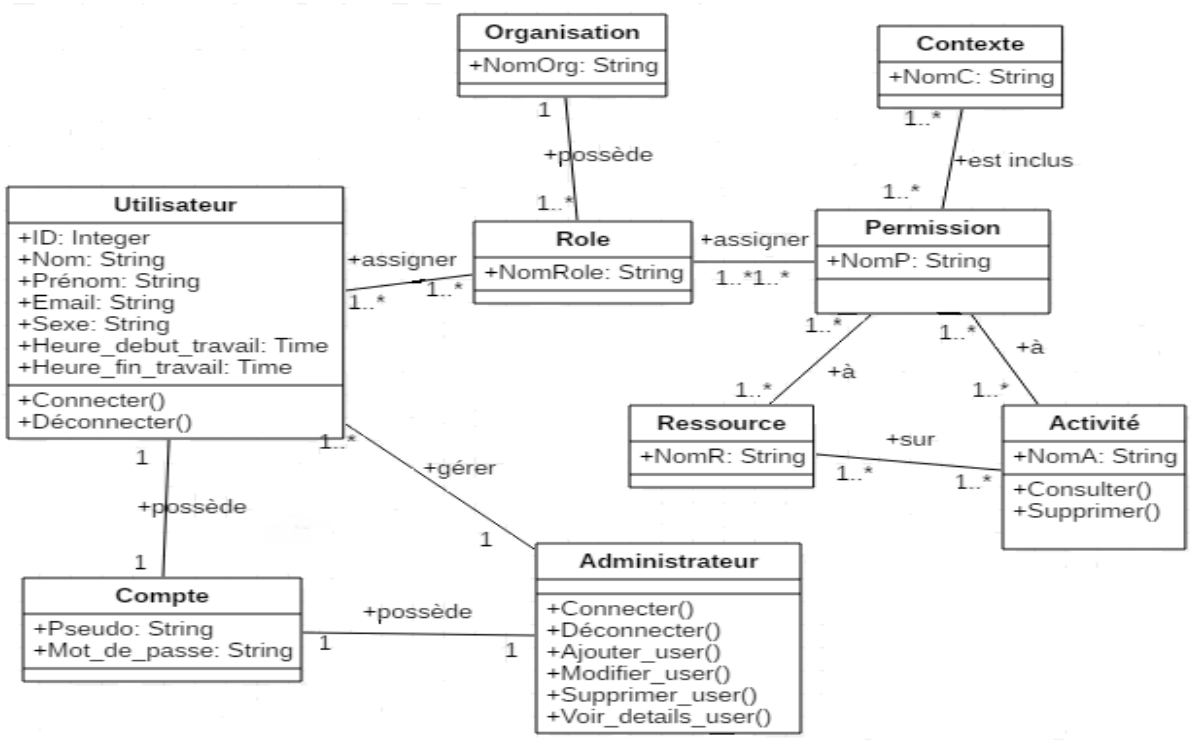

Figure 21 : Diagramme de classe

III.4 Conclusion

L'objectif de ce chapitre c'est de donner une idée sur la modélisation des risques en utilisant la méthode EBIOS au sein de l'organisation (MPTTN), passant à la modélisation de notre système par l'approche UML en présentant le diagramme de cas d'utilisation, les diagrammes de séquences et le diagramme de classe.

Dans le chapitre suivant, nous présenterons les étapes suivies dans la réalisation de notre application.

Chapitre IV : Implémentation

IV.1 Introduction

Dans le cycle de vie d'un logiciel, l'implémentation c'est la phase la plus importante après celle de la conception. Cette phase consiste à transformer le modèle conceptuelle vue précédemment en des composants logiciels formant notre système. Dans le présent chapitre nous présentons le système que nous avons conçu, ainsi que les outils et langages de programmation utilisés. Les principales interfaces du système sont montrées par des captures d'écran.

IV.2 Implémentation du système

IV.2.1 Description de fonctionnement du système

Le fonctionnement général du processus de contrôle d'accès de notre système est illustré par cette figure :

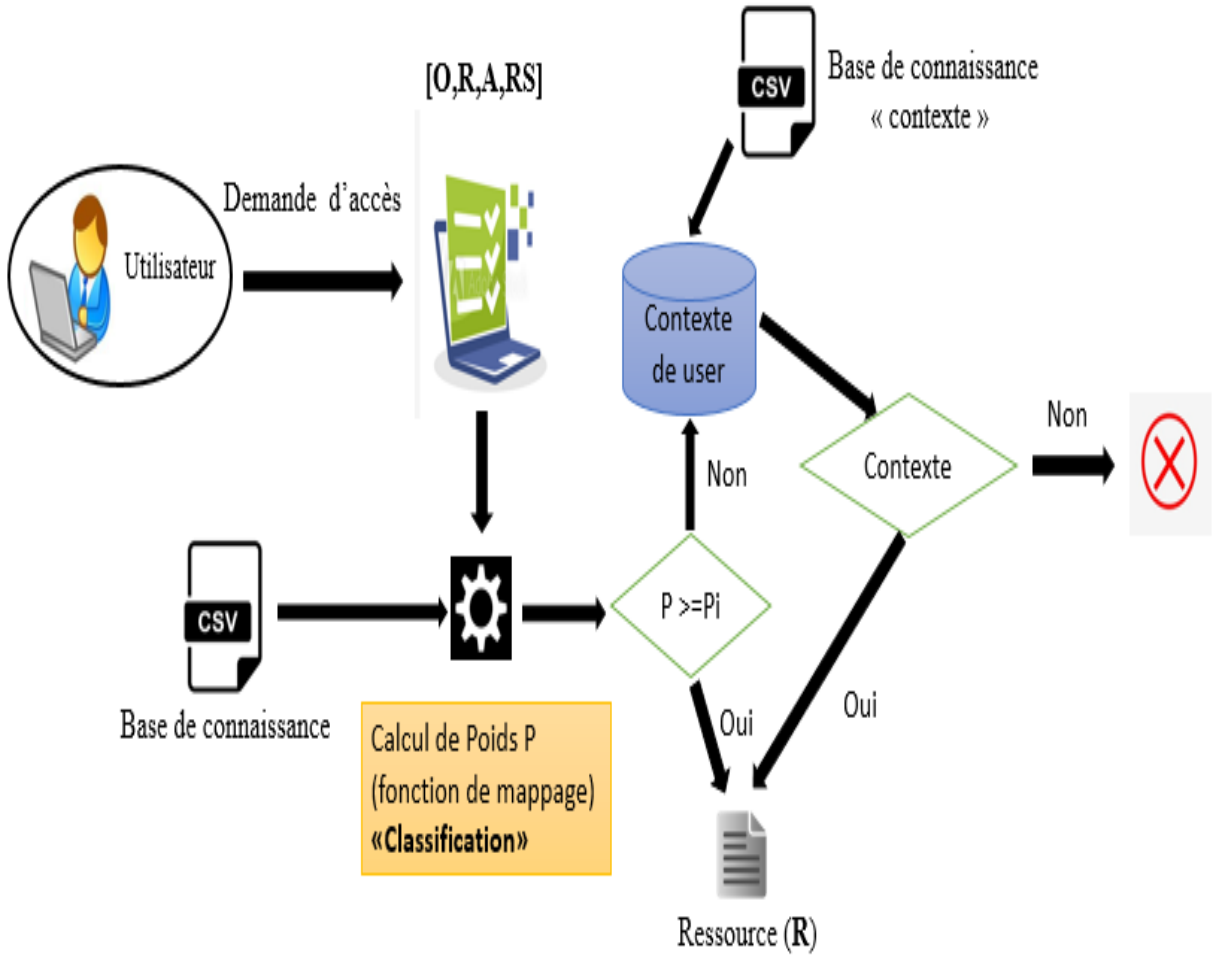

Figure 22 : Architecture du système

Lorsqu'un utilisateur fait une demande d'accès à une ressource, il doit remplir le formulaire d'accès qui contient les champs suivants : [organisation, rôle, activité, ressource]. Les entrées de chaque utilisateur représentent les attributs du modèle de contrôle d'accès ORBAC tel que : E1 = organisation, E2 = rôle, E3 = activité, E4 = ressource. Ces entrées seront traitées par un algorithme qui permet de calculer le poids pour chaque utilisateur en fonction de ses entrées. Cet algorithme prend en charge une base de connaissance sous forme d'un fichier csv qui contient les permissions de tous les utilisateurs (Figure 23).

	A	B	C	D
1	organisation,rôle,activité,ressource			
2	ptt,SDDSINA,consult,statistique			
3	ptt,RSSI,consult,analyse			
4	ptt,SDDSINA,delete,statistique			
5	ptt,RSSI,consult,fiche			
6	ptt,RSSI,consult,recrutement			
7	ptt,RSSI,delete,recrutement			
8	ptt,SDDSINA,consult,recrutement			
9	ptt,SDDSINA,consult,paiement			
10	ptt,SDDSINA,delete,paiement			
11	ptt,RSSI,consult,paiement			
12				

Figure 23 : Les permissions des utilisateurs

Chaque permission est représentée par une requête par rapport à ORBAC. Exemple de requête de l'utilisateur : **(organisation = ptt, rôle = RSSI, activité = consult, ressource = analyse)**. Après le calcul de poids, si le système trouve que le poids de l'utilisateur est supérieur ou égale au poids initial de la requête (4), l'utilisateur aura l'accès à la ressource avec l'activité demandée. Sinon, le système refuse l'accès de l'utilisateur.

Lorsque l'accès est refusé, le système vérifie si l'utilisateur a la permission de faire cette activité sur cette ressource dans un contexte donné (le contexte représente le cas d'urgence d'un utilisateur, c'est un attribut du modèle de contrôle d'accès ORBAC en plus des autres attributs cités précédemment. Dans notre cas, le contexte = 'urgence'). Le système demande à l'utilisateur de saisir son cas d'urgence (contexte). Ensuite, le contexte saisi sera comparé avec le contexte du système (l'ensemble des permissions des utilisateurs dans un contexte spécifique se trouvent dans un autre fichier csv similaire au fichier précédent (figure 24)). Si le système trouve que l'utilisateur existe dans ce contexte avec l'activité sur la ressource souhaitée, alors il donne l'accès à l'utilisateur, sinon il refuse l'accès de l'utilisateur définitivement.

	A	B	C	D
1	organisation,rôle,activité,ressource			
2	ptt,SDDSINA,consult,fiche			
3	ptt,RSSI,delete,analyse			
4	ptt,SDDSINA,delete,fiche			
5	ptt,RSSI,delete,recrutement			
6				
7				

Figure 24 : Les permissions des utilisateurs pour le contexte « urgence »

IV.2.2 Algorithme proposé basé sur l'apprentissage supervisé

Comme nous avons déjà cité, L'apprentissage supervisé consiste en des variables d'entrée (x) et une variable de sortie (Y), utiliser un algorithme pour apprendre la fonction de mappage de l'entrée à la sortie **Y=f (X)**. Le but est d'appréhender si bien la fonction de mappage que lorsque vous avez de nouvelles données d'entrée (x), vous pouvez prévoir les variables de sortie (Y) pour ces données. [31]

Dans notre cas, on propose cet algorithme qui permet de classer les utilisateurs (Accès ou Non accès) sans l'interaction de l'homme. Cette proposition ne demande pas une base de données ou une table des utilisateurs car notre système va se baser sur les poids des requêtes.

On a affecté à chaque entrée un poids et la somme des poids de ces entrées représente le poids de l'utilisateur. Le principe de l'algorithme est de calculer le poids et classer les utilisateurs.

- **Le principe de l'algorithme en général**

Inputs : Soit requête (organisation org, rôle R, activité A, ressource RS)
Pi= poids initial de la requête
Si E1 = org :
P [E1] = 1
Sinon P [E1] = 0
Si E2 = R :
P [E2] = 1
Si E4 = RS :
P [E4] = 1
Si E3 = A :
P [E3] = 1
Sinon P [E3] = 0
Sinon P [E4] = 0, P [E3] = 0
Sinon P [E2] = 0, P [E4] = 0, P[E3] = 0

Calculer la somme des poids
P = ∑ P [E i]
Comparer les P avec Pi
Si P ≥ Pi :
Accès = vrai
Sinon Accès = faux

- **Le principe de l'algorithme pour notre système**

Si E1 = 'ptt' :
P [E1] = 1
Sinon P [E1] = 0
Si E2 ∈ R :
P [E2] = 1
Si E4 ∈ RS :
P [E4] = 1
Si E3 ∈ A :
P [E3] = 1
Sinon P [E3] = 0
Sinon P [E4] = 0, P [E3] = 0
Sinon P [E2] = 0, P [E4] = 0, P [E3] = 0
Poids = P [E1] + P [E2] + P [E3] + P[E4]

Où :
{R : l'ensemble des rôles des utilisateurs de l'organisation}
{RS : l'ensemble des ressources existent au niveau de l'organisation}
{A : l'ensemble des activités possibles pour une ressource comme (delete, consult…)}
R, RS et A se trouvent au niveau de notre fichier csv (Figure 23).
Si le Poids >= Pi (4), l'utilisateur aura l'accès à la ressource avec l'activité demandée, sinon l'accès est refusé.

IV.2.3 Exemple des tests

Le tableau suivant montre les résultats des tests pour les différents utilisateurs sans établir un contexte spécifique (le cas normal) :

Organisation	Rôle	Activité	Ressource	Poids	Résultat
ptt (1)	RSSI (1)	consult (1)	analyse (1)	4	Accès
ptt (1)	SDDSINA (1)	delete (1)	statistique (1)	4	Accès
ptt (1)	RSSI (1)	delete (0)	analyse (1)	3	Non accès
ptt (1)	SDDSINA (1)	delete (0)	fiche (0)	2	Non accès
ptt (1)	DSI (0)	consult (0)	statistique (0)	1	Non accès

Tableau 11 : Exemple des tests

Le tableau suivant montre les résultats des tests pour les différents utilisateurs après l'utilisation de contexte (le contexte spécifique « urgence ») :

Organisation	Rôle	Activité	Ressource	Résultat
ptt	RSSI	delete	analyse	Accès
ptt	SDDSINA	consult	fiche	Accès
ptt	SDDSINA	delete	fiche	Accès
ptt	RSSI	delete	paiement	Non accès
ptt	RSSI	consult	statistique	Non accès

Tableau 12 : Exemple des tests après l'utilisation de contexte

D'après les résultats présentés dans les deux tableaux, nous remarquons que :

- Avec l'utilisation de concept « contexte » de ORBAC, il est possible d'avoir des utilisateurs avec d'autres permissions plus que les permissions présentes dans le cas normal. Par exemple, dans le cas normal le SDDSINA n'a pas d'accès au ressource 'fiche' pour une activité de 'consultation'. Mais avec l'établissement de contexte « urgence », cet utilisateur aura l'accès à cette ressource pour la même activité.
- Notre système donnera l'accès seulement aux utilisateurs autorisés qui sont présents dans les fichiers csv (figure 23, figure 24). Cela signifie que notre algorithme de classification donne des résultats performants.

IV.2.4 La cryptographie

Pour l'implémentation d'une fonction de cryptage on a utilisé la librairie OpenSSL pour le cryptage symétrique AES.

OpenSSL est une boite à outils de chiffrement comportant deux bibliothèques, libcrypto et libssl, fournissant respectivement une implémentation des algorithmes cryptographiques et du protocole de communication SSL/TLS, ainsi qu'une interface en ligne de commande. [61]

La figure suivante montre la fonction de cryptage implémentée :

```php
<?php

function convert_string($action, $string) {

  $output='';
  $encrypt_method='AES-256-CBC';
  $secret_key='eaiYYkYTysia2lnHiw0N0vx7t7a3kEJVLfbTKoQIx5o=';
  $secret_iv='eaiYYkYTysia2lnHiw0N0';
  $key=hash('sha256',$secret_key);
  $initialization_vector=substr(hash('sha256',$secret_iv),0,16);
  if($string != ''){
      if($action=='encrypt'){
          $output=openssl_encrypt($string, $encrypt_method, $key,0,$initialization_vector);
          $output=base64_encode($output);

      }

      if($action=='decrypt'){
          $output=base64_decode($string);
          $output=openssl_decrypt($output, $encrypt_method , $key,0,$initialization_vector);

      }
  }
   return $output;
}

?>
```

Figure 25 : La fonction de cryptage/décryptage de l'algorithme AES

Cette fonction prend deux paramètres :

- Le premier représente l'action, soit « encrypt » pour le cryptage ou « decrypt » pour le décryptage,
- Le deuxième représente le string que nous voulons crypter, dans notre cas c'est le mot de passe.

Le résultat après l'utilisation de cette fonction est montré dans la figure suivante :

id	nom	prenom	email	username	password	role	sexe	WHD	WHF	user_status
1	moussous	soumia	mous@gmail.com	soumia	UVpjbTFPazlGay9jK1Y4Sk94TkF6QT09	SDDSINA	Female	9	16	Active
2	boughelit	meriem	boug1@gmail.com	meriem	RjJSOXlkNjRlaXZBemlwQXFiWk9qdz09	RSSI	Female	9	16	Active

Figure 26 : Le résultat d'utilisation de fonction de cryptage pour le mot de passe

IV.3 Outils de développement

Dans cette partie, nous vous présentons l'ensemble des outils que nous avons utilisés pour mettre en œuvre la solution proposée.

IV.3.1 Outils matériels

Pour réaliser notre travail, on a disposé de :

- Un ordinateur portable « ACER » qui possède les caractéristiques suivantes :
  - Système d'exploitation : Windows 10 Professionnel 64 bits.
  - Processeur : Intel® Core™ i3-3217U.
  - CPU : 1.80 GHz.
  - Mémoire : 4 Go.
- Un ordinateur portable « LENOVO » qui possède les caractéristiques suivantes :
  - Système d'exploitation : Windows 8.1 64 bits.
  - Processeur : Intel® Pentium®.
  - CPU : 2,20 GHz.
  - Mémoire : 2Go.

IV.3.2 Outils logiciels

Les logiciels utilisés au cours de la réalisation de notre travail sont présentés comme suit :

- **PyCharm** : est un IDE (Integrated Development Environment). Il s'agit, tout comme jEdit, d'un logiciel permettant d'intégrer dans une même fenêtre tous les éléments utiles à la programmation en python : un éditeur de texte pour écrire des scripts, une console pour exécuter des programmes, ainsi qu'un explorateur de fichiers pour parcourir le projet en cours. [49]

- **XAMPP** : est un ensemble de logiciels servant à mettre en place aisément un serveur Web, un serveur FTP et un serveur de messagerie électronique. C'est une distribution de logiciels libres (X Apache MySQL Perl PHP) offrant une bonne souplesse d'utilisation, reconnue pour son installation simple et rapide. Ainsi, il est à la portée de la plupart des personnes dans la mesure où il ne requiert pas de connaissances spécifiques, et fonctionne plus sur les dispositifs d'exploitation les plus communs. [50]

- **Notepad++ :** Notepad++ est un éditeur de texte simple, qui peut-être une alternative à l'outil de Microsoft Notepad. Il permet de créer des textes simples, mais dispose des fonctions telles que la « coloration syntaxique » utile aux développeurs. [51]

- **StarUML :** est un logiciel de modélisation UML (Unified Modeling Language) open source qui peut remplacer dans bien des situations des logiciels commerciaux et coûteux comme Rational Rose1 ou Together2. Étant simple d'utilisation, nécessitant peu de ressources système, supportant UML 2, ce logiciel constitue une excellente option pour une familiarisation à la modélisation. Cependant, seule une version Windows est disponible. [52]

IV.3.3 Langages de programmation

Dans cette partie, nous allons présenter les langages de programmation utilisés pour réaliser notre projet.

- **Python3 :** Python est un langage de programmation puissant et facile à apprendre. Il dispose de structures de données de haut niveau et permet une approche simple mais efficace de la programmation orientée objet. Parce que sa syntaxe est élégante, que son typage est dynamique et qu'il est interprété, Python est un langage idéal pour l'écriture des scripts et le développement rapide d'applications dans de nombreux domaines et sur la plupart des plateformes. [53]

- **PHP :** est un langage informatique utilisé sur l'internet. Le terme PHP est un acronyme récursif de "PHP : Hypertext Preprocessor". Ce langage est principalement utilisé pour produire un site web dynamique. Il est courant que ce langage soit associé à une base de données, tel que MySQL. [54]

- **Html5 :** HyperText Markup Language 5, est une version du célèbre format HTML utilisé pour concevoir les sites Internet. Celui-ci se résume à un langage de balisage qui sert à l'écriture de l'hypertexte indispensable à la mise en forme d'une page Web. [55]

- **CSS** : Les feuilles de styles (en anglais "*Cascading Style Sheets*", abrégé CSS) sont un langage qui permet de gérer la présentation d'une page Web. Le langage CSS est une recommandation du World Wide Web Consortium (W3C), au même titre que HTML ou XML. Les styles permettent de définir des règles appliquées à un ou plusieurs documents HTML. Ces règles portent sur le positionnement des éléments, l'alignement, les polices de caractères, les couleurs, les marges et espacements, les bordures, les images de fond, etc. [56]

- **Javascript** : Est un langage de script orienté objet principalement utilisé dans les pages HTML. A l'opposé des langages serveurs (qui s'exécutent sur le site), Javascript est exécuté sur l'ordinateur de l'internaute par le navigateur lui-même. Ainsi, ce langage permet une interaction avec l'utilisateur en fonction de ses actions (lors du passage de la souris au-dessus d'un élément, du redimensionnement de la page...). [57]

- **Sql :** Le langage SQL (Structured Query Language) est un langage informatique utilisé pour exploiter des bases de données. Il permet de façon générale la définition, la manipulation et le contrôle de sécurité de données. [58]

- **JQuery** : est une bibliothèque JavaScript parmi les plus utilisées actuellement dans le développement web. Elle est utilisée dans le cadre de simple site web, RIA (Rich Internet Applications), jeux en HTML5, etc. Sa flexibilité permet en effet à jQuery de s'intégrer avec d'autres bibliothèques. [59]

- **Bootstrap :** est un Framework développé par l'équipe du réseau social Twitter. Proposé en open source (sous licence MIT), ce Framework utilisant les langages HTML, CSS et JavaScript fournit aux développeurs des outils pour créer un site facilement. Ce Framework est pensé pour développer des sites avec un design responsive, qui s'adapte à tout type d'écran, et en priorité pour les Smartphones. Il fournit des outils avec des styles déjà en place pour des typographies, des boutons, des interfaces de navigation et bien d'autres encore. On appelle ce type de Framework un "Front-End Framework". [60]

IV.4 Présentation de l'application

Au lancement du navigateur WEB, la page d'accueil suivante est présentée :

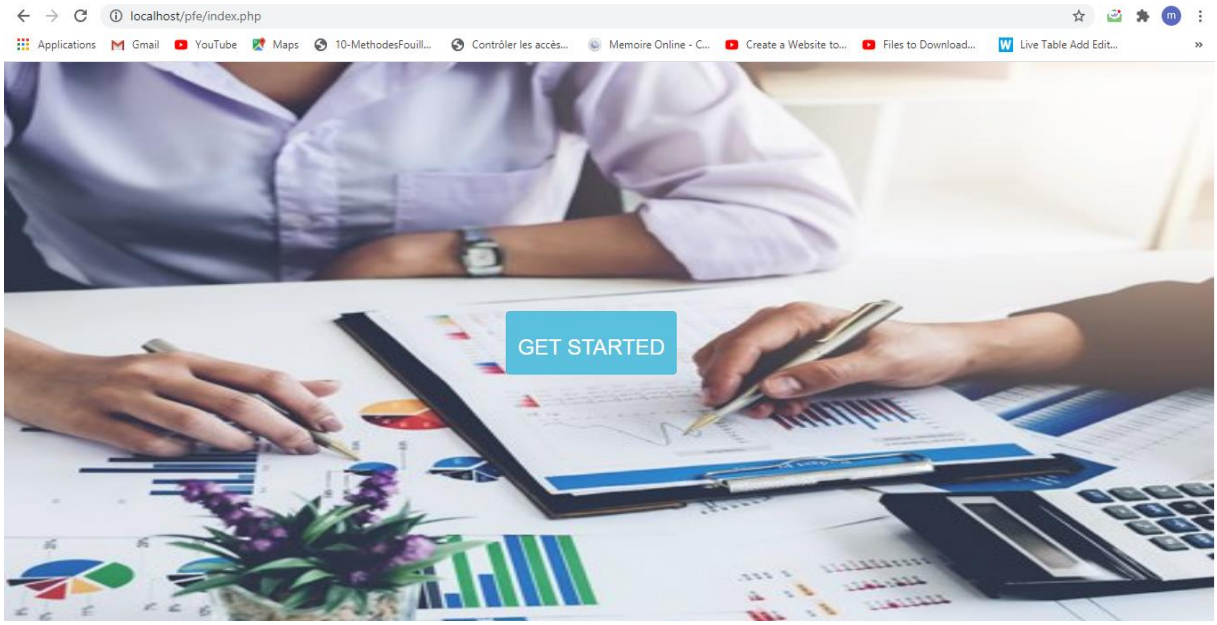

Figure 27 : L'interface principale de l'application

Tout d'abord il faut cliquer sur « GET STARTED » pour naviguer en toute sécurité. Alors qu'une page d'authentification est apparue. L'utilisateur doit saisir son Login et Mot de passe pour confirmer son identité.

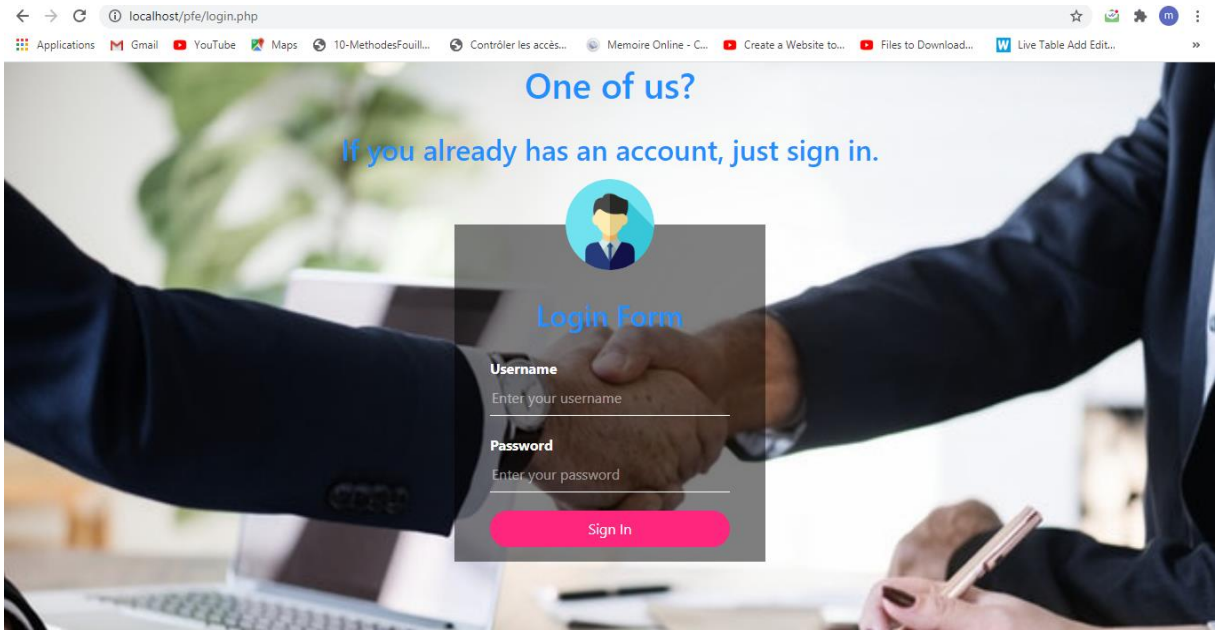

Figure 28 : Page d'authentification

Une fois l'identité est confirmée, l'utilisateur accède à son espace. Selon le type d'utilisateur (Administrateur, Employé), on aura deux espaces principaux :

IV.4.1 Espace administrateur

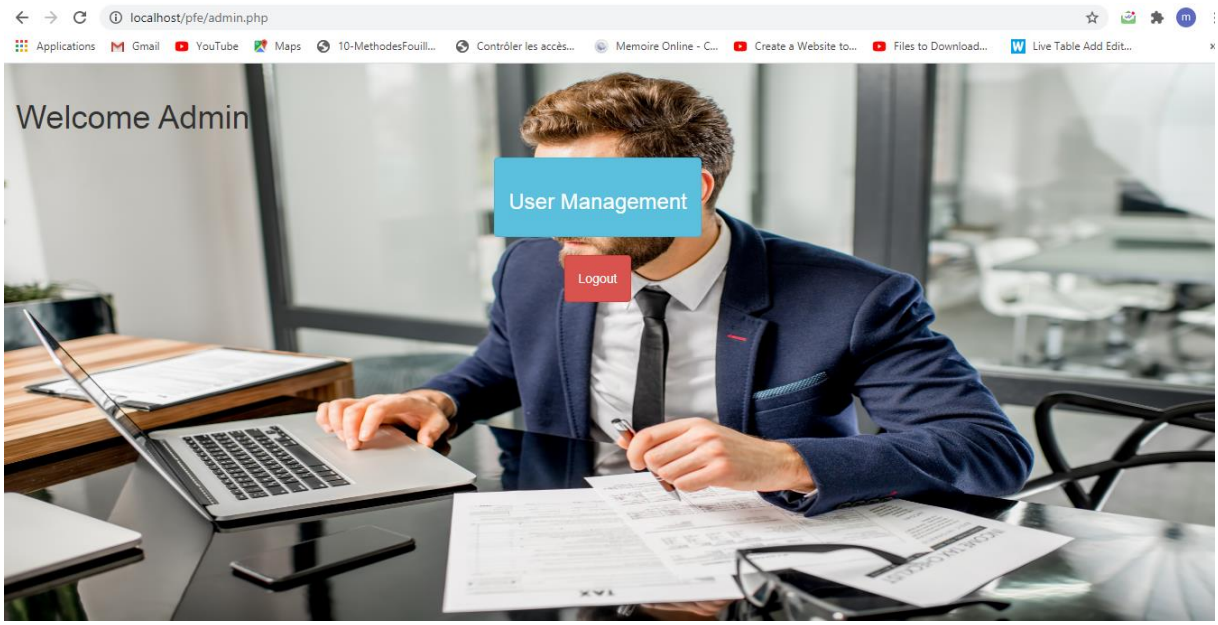

Figure 29 : L'interface principale de « Espace administrateur »

L'interface principale de l'administrateur est composée de deux champs :

- Le premier champ « User Management » c'est pour gérer les utilisateurs.
- Et le deuxième champ pour la déconnection.

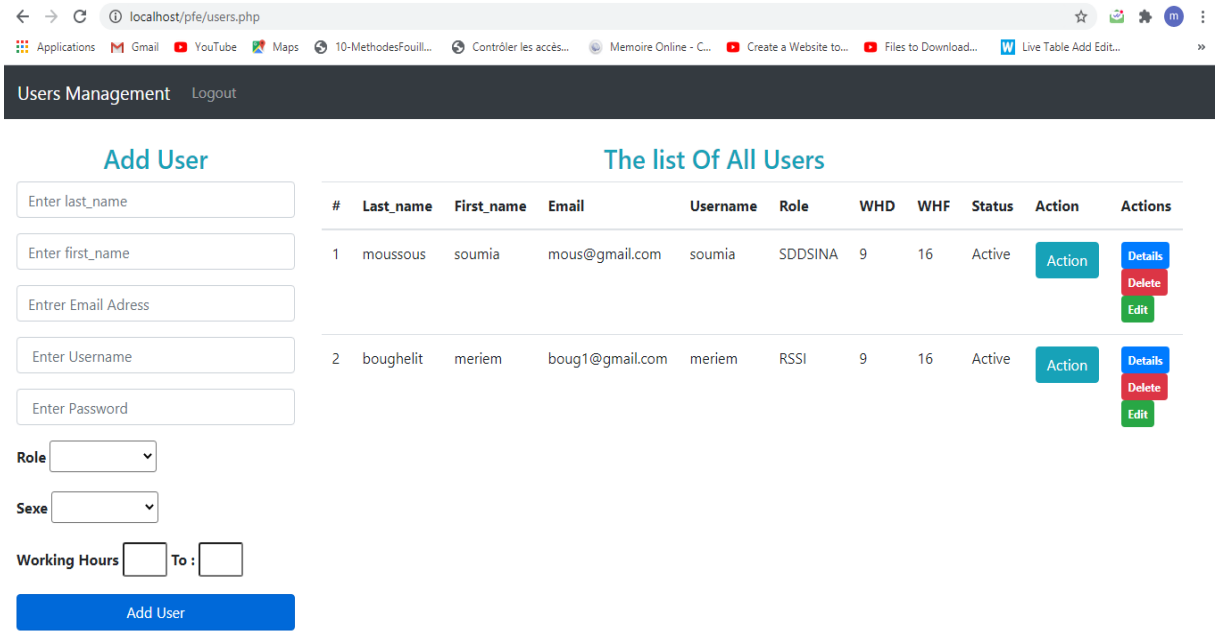

Figure 30 : Gestion d'utilisateur

Le rôle principal de l'administrateur est la gestion des utilisateurs. C'est-à-dire il a le droit d'ajouter un nouvel utilisateur et de lui fournir un login et un mot de passe pour qu'il puisse accéder à l'application. Ainsi, l'administrateur a le droit de modifier ou supprimer un utilisateur déjà créé et voir leurs détails. Il peut aussi désactiver les comptes des utilisateurs hors leurs heures de travail, et dans ce cas-là si un utilisateur essai de connecter, le système lui indique que son compte est désactivé par l'administrateur.

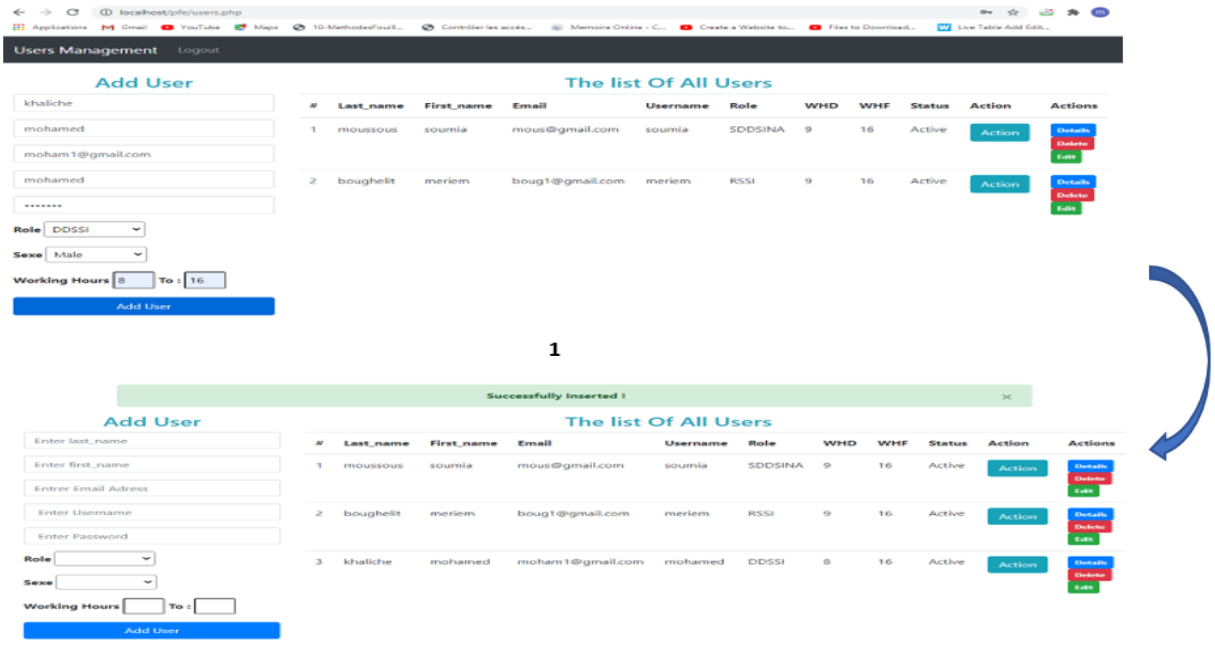

Figure 31 : Ajout d'un utilisateur

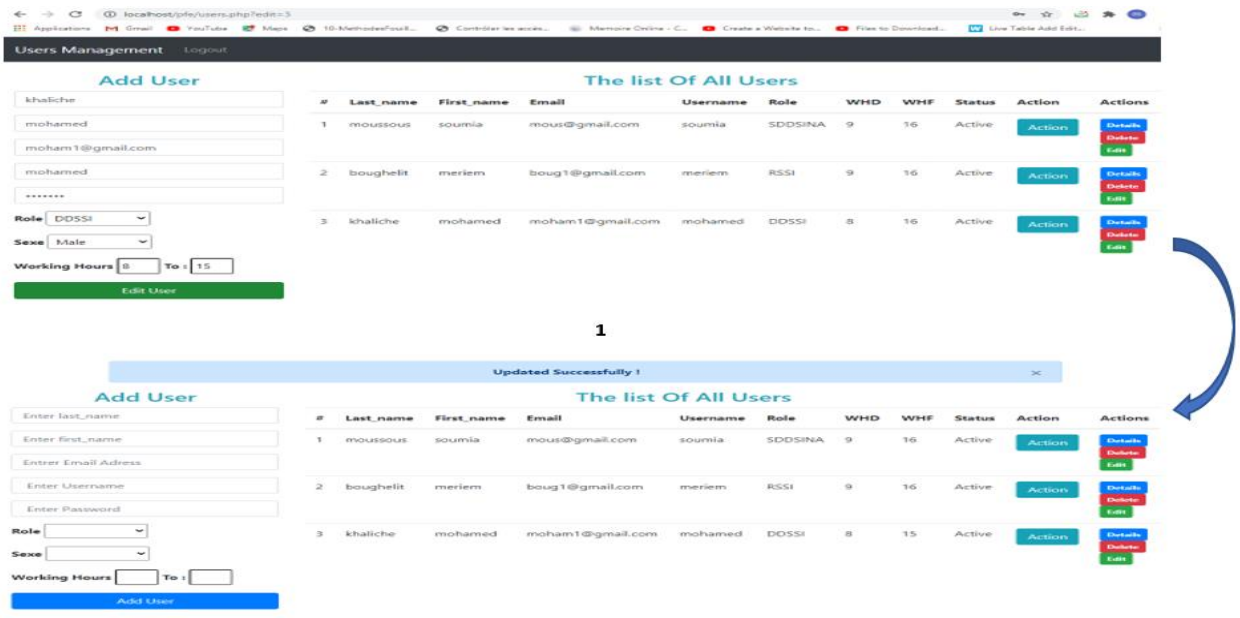

Figure 32 : Modification d'un utilisateur

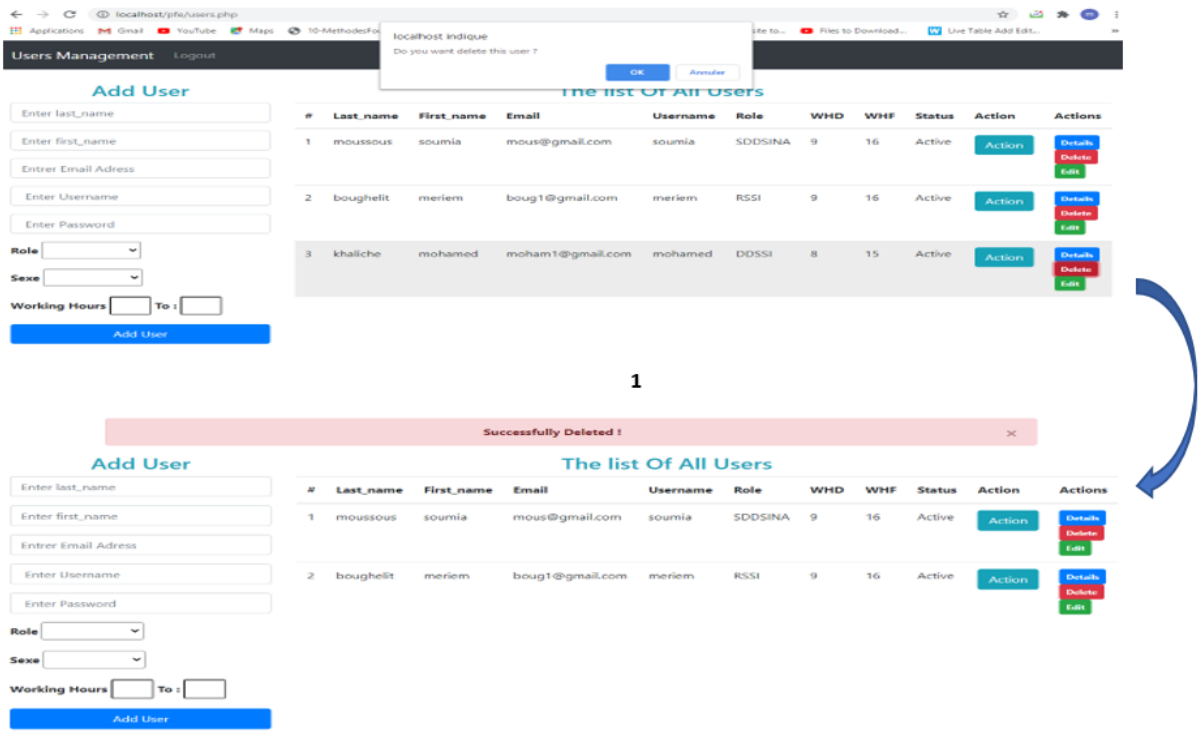

Figure 33 : Suppression d'un utilisateur

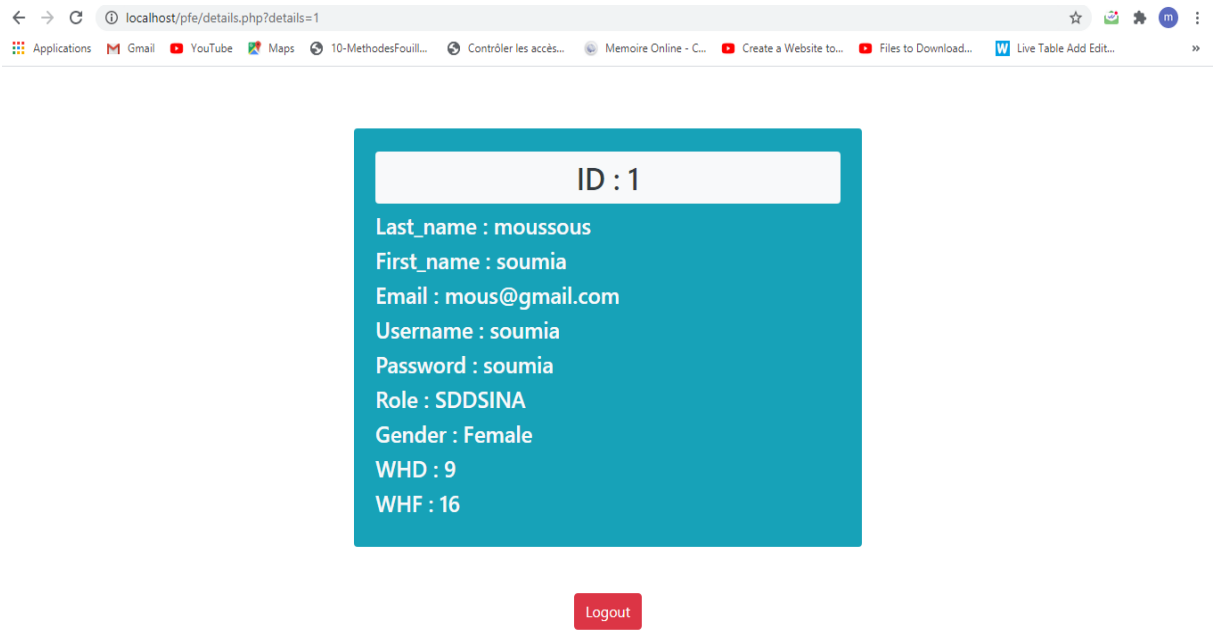

Figure 34 : Détails d'un utilisateur

IV.4.2 Espace utilisateur

Lorsqu'un utilisateur confirme son identité en passant par le formulaire d'authentification, un formulaire d'accès est apparu.

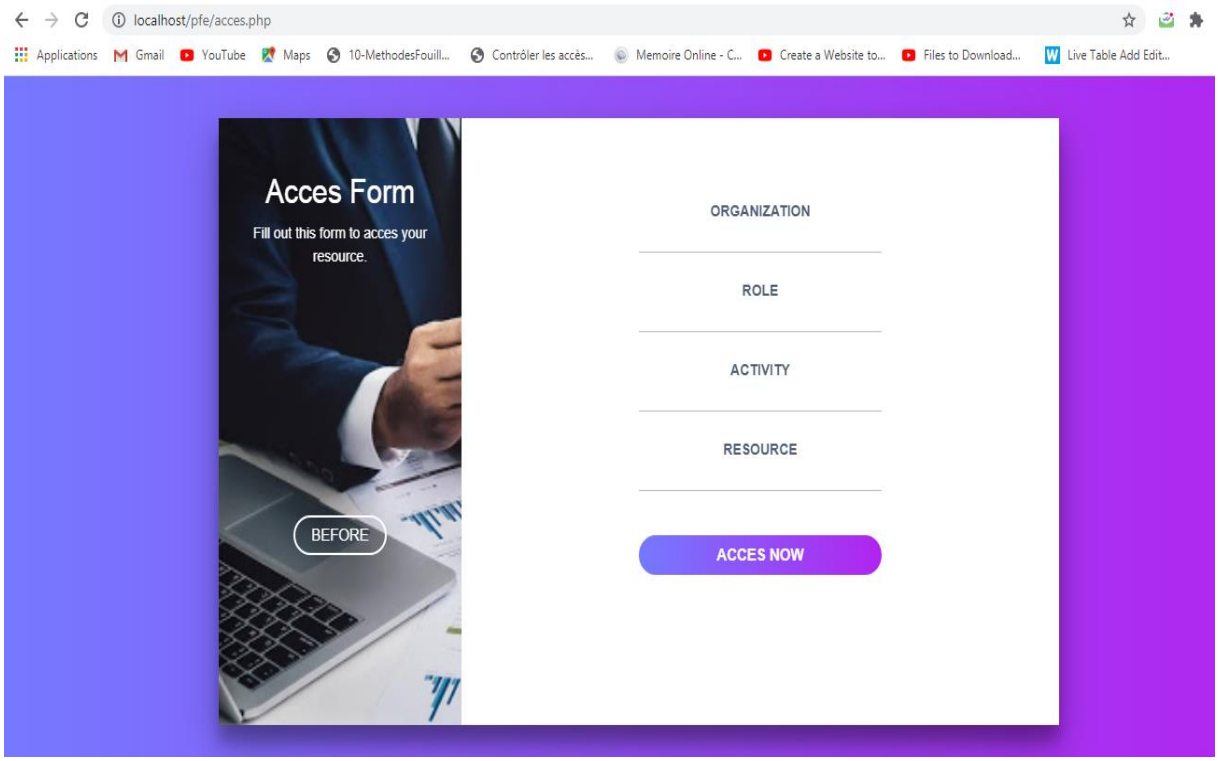

Figure 35 : Formulaire d'accès

Pour contrôler l'accès à la ressource demandée, l'utilisateur doit remplir tous les champs, ces champs sont les attributs du modèle de contrôle d'accès ORBAC :

- Le champ « ORGANIZATION » : signifie l'organisation dans laquelle l'utilisateur appartient, dans notre cas 'ptt'.
- Le champ « ROLE » : représente le rôle de l'utilisateur, exemple (RSSI : Responsable de Sécurité de Système d'Information).
- Le champ « ACTIVITY » : représente l'activité (delete, consult…).
- Le champ « RESOURCE » : représente la ressource sur laquelle l'activité est établie, exemple (le fichier statistique).

Après avoir terminé la saisie, le système contrôle l'accès de l'utilisateur. Si l'utilisateur saisie une organisation différente de 'ptt', alors un message d'erreur est présenté :

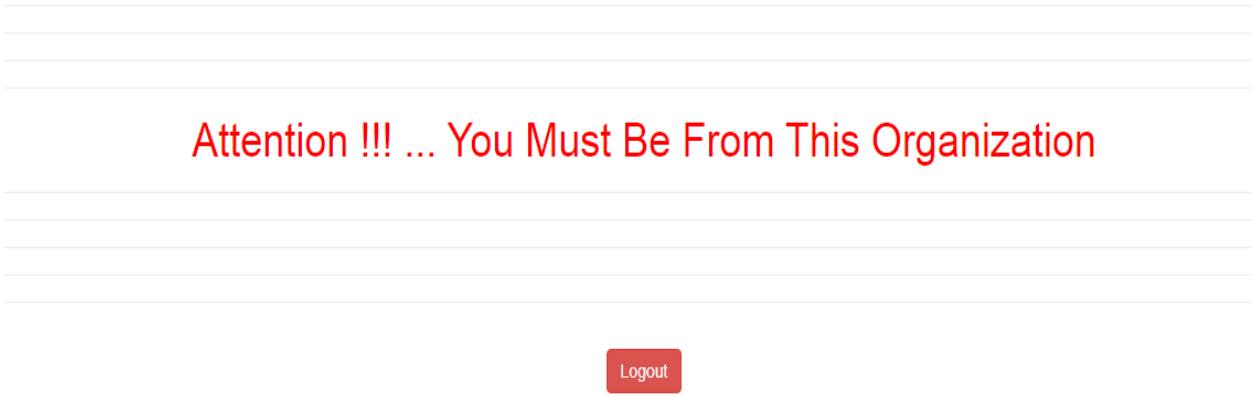

Figure 36 : Message d'erreur pour l'organisation

Si l'utilisateur a le droit de faire une activité sur une ressource, un espace contient la ressource demandée est présenté :

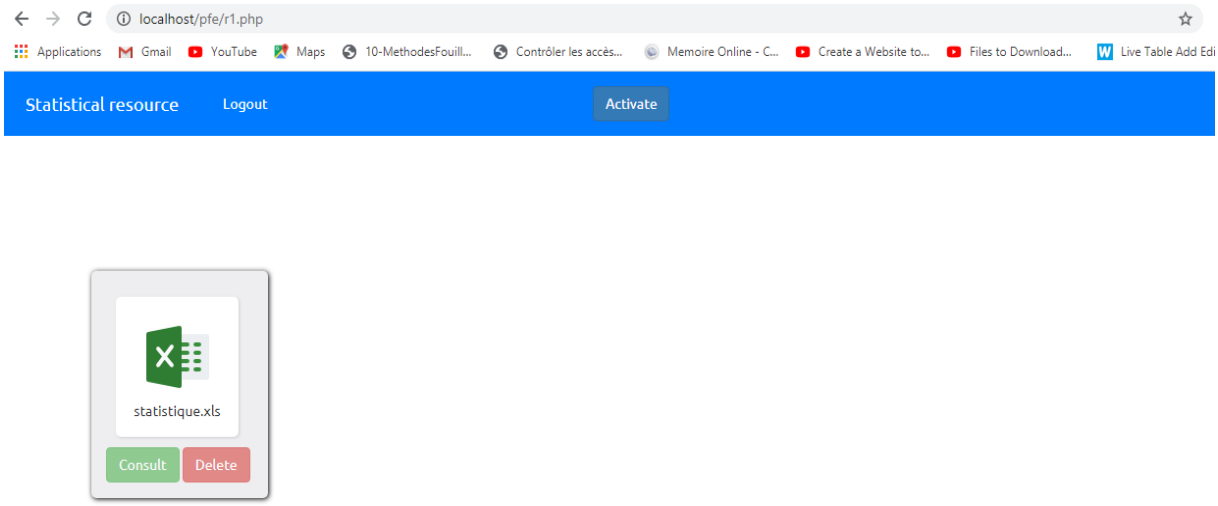

Figure 37 : Exemple d'une ressource

Au début les boutons (Consult et Delete) sont désactivés. Selon l'activité demandée, le bouton concerné est activé en cliquant sur « Activate ».

Si l'utilisateur n'a pas le droit de faire l'activité sur la ressource, un message d'erreur va apparaître et l'utilisateur doit citer son cas d'urgence.

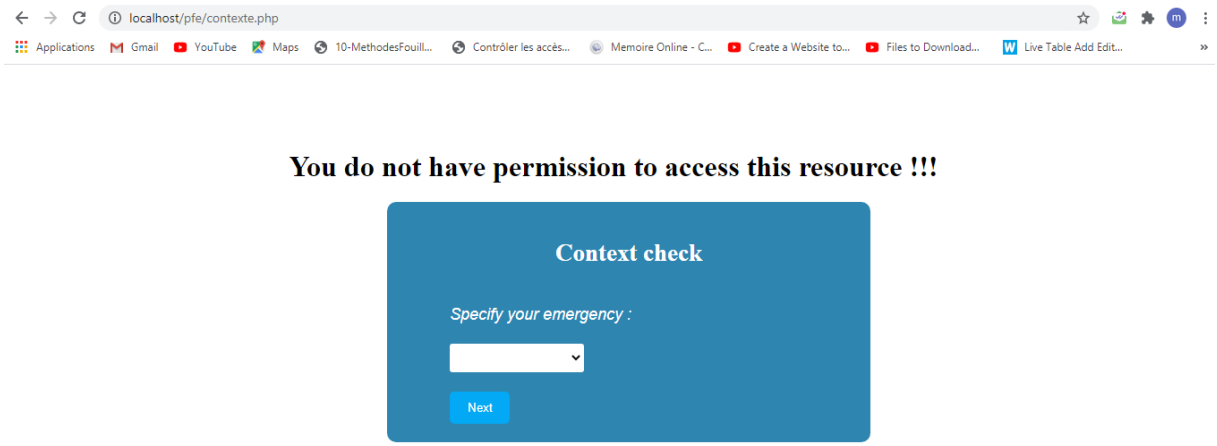

Figure 38 : Vérification de contexte

Le système ici vérifie si cet utilisateur a vraiment le droit de faire l'activité sur la ressource dans le contexte cité. Si oui, alors l'utilisateur est redirigé vers la page qui contient la ressource, sinon, le système refuse définitivement l'accès de l'utilisateur en affichant le message d'erreur suivant :

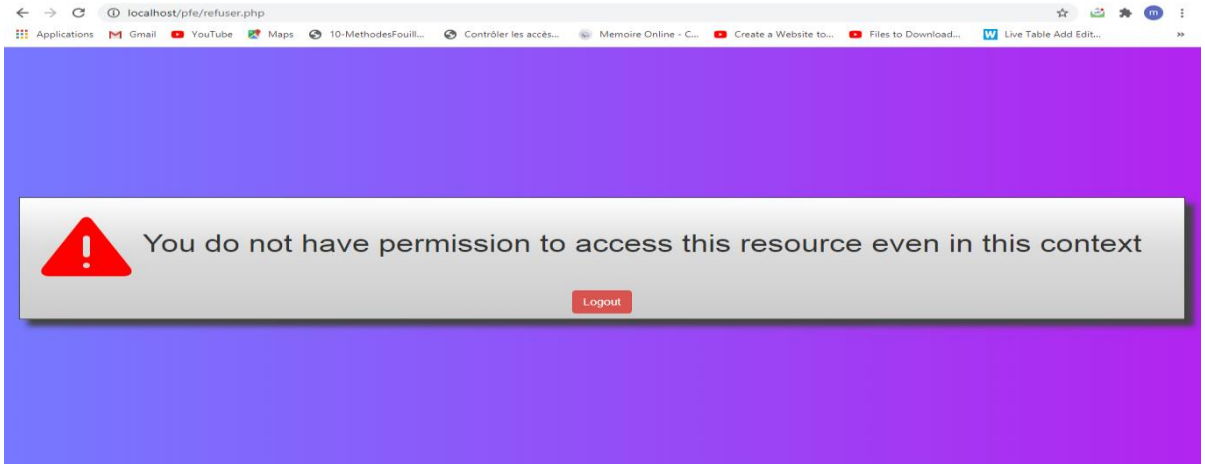

Figure 39 : Message de refus d'accès

IV.5 Conclusion

A ce chapitre, notre projet atteint sa fin. Au cours de ce dernier, nous avons décrit le processus de réalisation de notre application en présentant la méthode d'implémentation suivi par les différents outils de développement nécessaires à la mise en place de notre système. Par la suite, nous avons présenté les principales interfaces à travers lesquels les utilisateurs peuvent accéder aux différentes fonctionnalités comme la gestion des utilisateurs pour l'administrateur et la gestion d'accès à la ressource pour les utilisateurs (employés).

Conclusion générale

En conclusion de notre travail de fin d'étude, notre objectif est de développer un système d'accès intelligent, tout en assurant la sécurité d'accès à l'information. Ce système est basé sur le principe de « l'apprentissage supervisé et le contrôle d'accès basé sur l'organisation (ORBAC) » et ce, pour sortir de l'ancienne manière de l'accès traditionnel.

Afin de vous permettre une vision globale sur notre objectif de sécurité, nous avons préalablement présenté d'abord la sécurité informatique et ces objectifs, ainsi que, quelques attaques, logiciels malveillants et techniques de défense. Puis nous avons présenté le contrôle d'accès et ces différents modèles.

Ensuite et dans la même voix, nous avons présenté l'apprentissage automatique, ces principes, ainsi que, ces types et quelques algorithmes les plus connus (deuxième chapitre).

Par la suite dans le troisième chapitre, on fait une modélisation des risques avec la méthode EBIOS au sein du MPTTN, puis on a passé à la modélisation de notre projet en utilisant l'approche UML dans laquelle on a expliqué le fonctionnement de notre solution.

Nous avons terminé cette mémoire par l'implémentation de la solution proposée pour la sous-direction SSDDNA, qui est un portail web pour gérer les accès à la ressource demandée, où chaque employé qui veut accéder à certaines informations doit passer par le formulaire d'accès, qui est liée par une base de connaissance contienne les permissions de chaque employé. Avec cette solution on est arrivé à garantir la confidentialité, l'intégrité et la disponibilité des ressources de l'organisation.

Enfin, nous proposons comme perspectives :

- Concernant les activités sur les ressources : Ajouter d'autres activités comme transférer des fichiers par exemple (dans notre travail on a seulement consultation et suppression).
- Concernant les ressources : Elargir l'intervalle des autorisations possibles avec d'autres types de ressources.
- Pour la cryptographie : utiliser des mécanismes cryptographiques de haut niveau sur l'ensemble des données du système.
- Notre solution est un exemple effectué dans la sous-direction SSDDNA, mais nous avons les facultés de l'élargir à l'ensemble de l'organisation.

Bibliographie

[1] AINENNAS Faiza, ZIDI Nassima, **« contrôle d'accès aux services sensibles au contexte »**, Mémoire de Master en Informatique, Université Abderahman Mira de Béjaia ,2015, pp (16-33).

[2] BIR Khaled, SAOUDI Yanis, **« Mise en place d'un système de détection d'intrusion »**, Mémoire de Master Professionnel en Informatique, Université Abderrahmane Mira Béjaïa, 2017, pp (7-12).

[3] BOUKHLOUF Djemaa, « **Une approche à base d'agents mobiles pour la sécurité des systèmes d'informations sur le web »**, Thèse de Doctorat en Sciences en Informatique, Université Mohamed Khider Biskra, 2016, pp (19-25).

[4] M. TOUATI Azzedine, « **Détection d'intrusions dans les réseaux LAN : Installation et configuration de l'IDS-SNORT »**, Mémoire de Master professionnel en Informatique, Université A/Mira de Béjaïa, 2016, pp (3-15).

[5] BELKHATMI Keltouma, BENAMARA Ouarda, « **Mise en place d'un système de détection et de prévention d'intrusion »**, Master en Informatique, Université A/Mira de Béjaïa, 2016, pp (21-23).

[6] landry Ndjate, **« mise en place d'un crypto système pour la sécurité des données et la détection d'intrusion dans un supermarché »**, Mémoire en ligne, Université Notre Dame du Kasayi – Graduat, 2014, disponible sur : https://www.memoireonline.com/01/16/9388/m_Mise-en-place-dun-crypto-systeme-pour-la-securite-des-donnee-et-la-detection-dintrusion-da7.html.

[7] Laurent Bloch, Christophe Wolfhugel, **« Sécurité informatique »**, EYROLLES, 2eme edition, 2005.

[8] Worachet UTTHA, **« Étude des politiques de sécurité pour les applications distribuées : le problème des dépendances transitives »**, Thèse de Doctorat Informatique, Université d'Aix-Marseille, 2016, pp (23-24).

[9] Sofiene Boulares, **« Validation des politiques de sécurité par rapport aux modèles de contrôle d'accès »**, Mémoire de Magister en Informatique, Université du Québec en Outaouais, Aout 2010, pp (11-33).

[10] KHELIFA Nor Eddine, **« Intégration du modèle de contrôle d'accès RBAC (Role-Based Access control) dans les diagrammes UML (Cas d'Utilisation et Séquence »**, Mémoire de Magister Informatique, université d'Oran, pp (18-19).

[11] Sourour JEMILI, « **ANALYSE DE RISQUE DANS LES SYSTEMES DE CONTRÔLE D'ACCES »,** Mémoire de Magister en Informatique, Université du Québec en Outaouais, 2013, p14.

[12] DEBIANE noureddine, ZEGMALI fatima zohra, « **Développement d'un modèle pour le contrôle d'accès au dossier médical personnel (Etude de cas : CHU Algérien)** », Mémoire de Master en Genie Biomedical, Université Abou Bakr Belkaïd de Tlemcen, 2017, pp (24-25).

[13] https://fr.wikipedia.org/wiki/Controle_d'accès_basé_sur_l'organisation, consulté le [03/05/2020].

[14] Mahamat Ahmat Abakar, **« Étude et mise en œuvre d'une architecture pour l'authentification et la gestion de documents numériques certifiés Application dans le contexte des services en ligne pour le grand public »**, Thèse de Doctorat Génie Informatique, université Jean Monnet de Saint-Étienne, 2012, p17.

[15] Anas ABOU EL KALAM, **« MODÈLES ET POLITIQUES DE SECURITE POUR LES DOMAINES DE LA SANTE ET DES AFFAIRES SOCIALES »**, Thèse de Doctorat Informatique, Institut National Polytechnique de Toulouse, 2003, p100.

[16] DERBALE Asma et MEHENNI Ouassila **« Sécurisation du dossier médical à base d'ontologie »**, Mémoire de Master en Genie Biomedical, Université Abou Bakr Belkaïd de Tlemcen, 2019, pp (15-17).

[17] Anas Abou El Kalam, Rania El Baida, Philippe Balbiani, Salem Benferhat, Frédéric Cuppens, Yves Deswarte, Alexandre Miège, Claire Saure, Gilles Trouessin. **« ORBAC : un modèle de contrôle d'accès basé sur les organisations »**, article, pp (3-4).

[18] Amine BAINA, « **Contrôle d'Accès pour les Grandes Infrastructures Critiques : Application au réseau d'énergie électrique »**, Thèse de Doctorat Systèmes Informatiques Critiques, Université de Toulouse, 2009, pp (49-52).

[19] Carine Arlette FOTSO TAGNE, **« Conception et vérification de la cohérence d'une politique de sécurité dans un réseau local »**, Mémoire en ligne, Université de Yaoundé 1,2013, p6, disponible sur : https://www.memoireonline.com/11/15/9285/Conception-et-verification-de-la-coherence-dune-politique-de-securite-dans-un-reseau-local.html.

[20] Aymen Kamoun, **« Adaptation d'architectures logicielles de contrôle d'accès dans les environnements collaboratifs ubiquitaires »**, Thèse de Doctorat Informatique, Université de Toulouse, 2014, p22.

[21] BENMANSOUR Adel, BOUZOUINA Sara Safia, OUDINAT Sara Hadjer, **« L'aide de l'intelligence artificielle à la prise de décision (Cas d'étude : Classification des tubes PEHD de la société Canal Plast Groupe Kherbouche) »**, Mémoire de Master Génie Industrielle, Université Abou Bekr Belkaid – Tlemcen, 2017, pp (23-75).

[22] Zara Islem, **« L'intelligence artificielle principe, outils et objectifs. »**, Mémoire de Master Automatique, Université Badji Mokhtar Annaba, 2019, pp (12-23).

[23] https://blog.lukas.fr/lintelligence-artificielle-machine-learning-deep-learning, consulté le [17/04/2020].

[24] https://le-datascientist.fr/apprentissage-supervise-vs-non-supervise , consulté le [10/05/2020].

[25] HAMMOUD Djamila, **« Apprentissage Automatique dans un Agent »**, Thèse de Doctorat Informatique, Université Mentouri Constantine, p34.

[26] Jiawei Han, Micheline Kamber, Jian Pei, **« Data mining: Concepts and Techniques »** 3rd Edition, Elsevier, USA, 2012, p18.

[27] Bruno Agard, Andrew Kusiak, **« EXPLORATION DES BASES DE DONNÉES INDUSTRIELLES À L'AIDE DU DATA MINING – PERSPECTIVES »**, 9ème Colloque National AIP PRIMECA, avril 2005, p3.

[28] Daniel T. Larose, Chantal D. Larose, **« Data Mining and Predictive Analytics »** Second Edition, John Wiley & sons, Canada, 2015.

[29] Minimol Anil Job**, « Data Mining Techniques Applying on Educational Dataset to Evaluate Learner Performance Using Cluster Analysis »**, EJERS, European Journal of Engineering Research and Science Vol. 3, No. 11, November 2018, p26.

[30] Jiawei Han, Micheline Kamber, Jian Pei, **«Data Mining: Concepts and Techniques »**, 3rd Edition Solution Manual, Morgan Kaufmann, 2011, p5.

[31] https://waytolearnx.com/2018/11/difference-entre-apprentissage-supervise-et-non-supervise.html , consulté le [23/03/2020].

[32] https://openclassrooms.com/fr/courses/4011851-initiez-vous-au-machine-learning/4020611-identifiez-les-differents-types-dapprentissage-automatiques , consulté le [01/04/2019].

[33] CHIKOUCHE Soumia, « **Système de détection d'intrusion basé sur la classification comportementale des processus »**, Mémoire de Master en Informatique, Université de M'sila, 2012, pp (64-66).

[34] https://mrmint.fr/lapprentissage-non-supervise-machine-learning, consulté le [29/05/2020].

[35] LABIAD ALI, **« SÉLECTION DES MOTS CLÉS BASÉE SUR LA CLASSIFICATION ET L'EXTRACTION DES RÈGLES D'ASSOCIATION »**, Mémoire de Magister en Mathématiques et Informatique Appliquées, Université du Québec À Trois-Rivières, 2017, pp (19-27).

[36] ZAHRA YAHIAOUI, **« Etude et implémentation de l'algorithme c moyenne floue et ses variantes »**, Mémoire de Master, Université de M'sila, 2013, pp (35-36).

[37] Françoise Fessant, **« Apprentissage non supervisé »**, TECH/SUSI, 28/09/2006, p12, disponible sur : http://www.vincentlemaire-labs.fr/cours/2.2-ApprentissageNonSupervise.pdf.

[38] HAMIDOUCHE Saddek, IDJERAOUI Tayeb, **« Clustering : Approche par la théorie des jeux »**, Mémoire de Master en Recherche Opérationnelle, Université Mira Abderrahmane de Bejaïa, 2013, pp (19-20).

[39] https://le-datascientist.fr/les-svm-support-vector-machine , consulté le [20/02/2019].

[40] GHALI Ahmed, **« Amélioration de la reconnaissance par le visage »**, Mémoire de Magister en Informatique, Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf, 2015, p19.

[41] https://www.researchgate.net/figure/Exemple-de-classification-binaire-lineaire-et-non-lineaire-dans-R-2-a-probleme_fig12_330441297, consulté le [02/06/2019].

[42] https://medium.com/@gifadelyaninursyafitri/classification-with-knn-using-r-fa4b3a7eea3c, consulté le [04/11/2018].

[43] ZAIZ Faouzi, **« Les Supports Vecteurs Machines (SVM) pour la reconnaissance des caractères manuscrits arabes »**, Mémoire de Magister en Informatique, Université Mohamed Khider – Biskra, 2010, p17.

[44] Pascal Bou Nassar, **« Gestion de la sécurité dans une infrastructure de services dynamique : Une approche par gestion des risques »**, Le grade de docteur, Formation doctorale : Informatique – Gestion de la sécurité, INSA de Lyon, 2012, pp (77-79), disponible sur : http://theses.insa-lyon.fr/publication/2012ISAL0102/these.pdf.

[45] BORDJAH Dahia, BOUDJADI Amel, **« Audit et définition d'une politique de sécurité. Cas d'étude SONATRACH DP. »**, Mémoire de Master en Informatique, Université Abderrahmane Mira de Béjaïa, 2013, pp (36-44).

[46] MIROUD Mohammed El Mustapha, **« La sécurité dans les systèmes de e-santé »**, Mémoire de Magister en Informatique, Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf, 2016, pp (57-58).

[47] **« Expression des Besoins et Identification des Objectifs de Sécurité »**, EBIOS-ÉTUDE DE CAS @RCHIMED, Paris, Version du 25 janvier 2010, p16, disponible sur : https://www.ssi.gouv.fr/uploads/2011/10/EBIOS-EtudeDeCas-Archimed-2010-01-25.pdf.

[48] https://openclassrooms.com/fr/courses/2035826-debutez-lanalyse-logicielle-avec-uml, consulté le [03/02/2020].

[49] https://doplab.unil.ch/wp-content/uploads/2018/09/Pycharm_Tutorial-v1.pdf, consulté le [01/09/2020].

[50] http://www.standard-du-web.com/xampp.php , consulté le [27/08/2020].

[51] https://www.pack-logiciels-libres.fr/spip.php?logiciel34 , consulté le [27/08/2020].

[52] https://inf1410.teluq.ca/teluqDownload.php?file=2014/01/INF1410 PresentationStarUML.pdf, consulté le [29/08/2020].

[53] https://docs.python.org/fr/3/tutorial/ , consulté le [27/08/2020].

[54] http://glossaire.infowebmaster.fr/php/ , consulté le [27/08/2020].

[55] https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203257-html5-hypertext-markup-langage5-definition-traduction/ , consulté le [29/08/2020].

[56] https://www.futura-sciences.com/tech/definitions/internet-css-4050/ , consulté le [27/08/2020].

[57] https://www.futura-sciences.com/tech/definitions/internet-javascript-509/, consulté le [29/08/2020].

[58] https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203603-sql-structured-query-language-definition-traduction-et-acteurs, consulté le [29/08/2020].

[59] http://tecfaetu.unige.ch/etu-maltt/tetris/fritz0/stic-2/ex16/introduction-jquery/, consulté le [29/08/2020].

[60] https://www.journaldunet.com/web-tech/developpeur/1159810-bootstrap-definition-tutoriels-astuces-pratiques/ , consulté le [29/08/2020].

[61] https://fr.wikipedia.org/wiki/OpenSSL, consulté le [13/09/2020].