République Algérienne Démocratique et Populaire

Ministère de l'enseignement Supérieur et de la Recherche scientifique

Université Saad Dahleb Blida 1

Faculté des Science et de la technologie

Département d'Informatique

Mémoire de fin d'étude

Pour l'obtention d'un diplôme de Master en Informatique

Thème

Plateforme de simulation des tests d'intrusion

Réalisé par

- AICHE Mohamed — **Spécialité** : SIR
- TEKABDJI Youcef — **Spécialité** : IL

Soutenu devant le jury composé de

Mme Boutoumi Bachira	Présidente
Mme Arkam Meriem	Examinatrice
Mme Boustia Narhimene	Promotrice

Organisme d'accueil : Intervalle Technologies

Année Universitaire : 2019 - 2020

Remerciements

Nous commençons par remercier et rendre grâce à ALLAH le tout puissant de nous avoir permis et nous avoir donné les moyens de mener à bon terme ce travail.

Nous tenons également à remercier l'équipe d'Intervalle Technologies pour l'accueil qu'elle nous a préservé et plus précisément à Messieurs Abdelmoumen Benamarouch et Adda Mohamed pour leur soutien, les informations qu'ils nous ont accordé, les références, les réflexions et pour leur temps qui nous ont accordé.

Nous remercions également nos professeurs pour la qualité de l'enseignement qu'ils nous ont mis à notre disposition au cours des cinq années d'études, nous tenons à remercier plus particulièrement Mme Boustia Narhimene pour son encadrement, son orientation, ses conseils et sa disponibilité qui nous ont permis de mener bien ce travail.

Pour finir, Nous remercions le président et les membres du jury pour avoir accepté d'examiner notre travail.

Dédicaces

Je dédié ce travail a :

A tous ceux qui se sont sacrifiés pour nous offrir les conditions propices à notre réussite :

A ma mère, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, pour le sens du devoir qu'elle m'a enseigné depuis mon enfance.

A mon frère et ma sœur qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

A tous mes amis et mes proches pour leurs encouragements et leurs soutiens inconditionnels.

A. Mohamed

Je dédié ce travail a :

A MA TRÈS CHÈRE MÈRE

Autant de phrases aussi expressives soient-elles ne sauraient montrer le degré d'amour et d'affection que j'éprouve pour toi. Tu m'as comblé avec ta tendresse et affection tout au long de ma vie. Tu n'as cessé de me soutenir et de m'encourager durant toutes les années de mes études, tu as toujours été présente à mes côtés pour me consoler quand il fallait. En ce jour mémorable, pour moi ainsi que pour toi, reçoit cet effort en signe de ma vive reconnaissance et de ma profonde estime et respect. Puisse le tout puissant te donner santé, bonheur et longue vie afin que je puisse te combler à mon tour. Maman, je te serai toujours redevable.

T. Youcef

Résumé

A l’heure actuelle on peut constater que l'informatique est devenue un moyen indispensable et incontournable dans la vie quotidienne de l’homme.

L’utilisation d’un nombre innombrable d’applications et de logiciels qui sont toujours en cours de développement actuellement, nécessite la sécurisation des données et la sécurité de ce monde.

L’objectif de notre travail consiste à concevoir une plateforme, cette dernière est basée sur plusieurs challenges et simulations qui aident les amateurs de la sécurité informatique, les employées des entreprises Et les clients de se familiariser des différents problèmes de sécurité avec une assistance intelligente.

Mots clés: sécurisation des données, challenges, simulations, sécurité informatique, assistance intelligente.

Abstract

Today we can see that computing has become an essential and unavoidable means in human daily life.

The use of the countless number of applications and software that are still under development today, requires the security of data and the security of this world.

The objective of our work is to design a platform, the latter is based on several challenges and simulations that are aimed at informatics security enthusiasts, company employees and customers to familiarize themselves with the various security issues with an intelligent assistance.

Keywords: security of data, challenges, simulations, Informatics security, intelligent assistance.

ملخص

يمكننا اليوم أن نرى أن مجال الإعلام الآلي أصبح وسيلة لا غنى عنها ولا مفر منها في حياة الإنسان اليومية.

يتطلب استخدام العدد الكبير الذي لا يمكن عده من التطبيقات والبرامج التي لا تزال قيد التطوير والبرمجة حتى اليوم تأمين البيانات وتأمين هذا العالم.

الهدف من عملنا هو تصميم منصة إلكترونية، ويستند هذا الأخير إلى العديد من التحديات في مجال الأمن والمحاكاة التي تستهدف المتحمسين للأمن المعلوماتي وموظفي الشركة والعملاء للتعرف على مشاكل الأمن المعلوماتي المختلفة مع الاعتماد على المساعدة ذكية.

الكلمات المفاتيح: تأمين البيانات، التحديات، المحاكاة، الأمن المعلوماتي، مساعدة ذكية.

Table des matières

Table des Figures

Glossaire

ANN	Artificial Neural Networks
APT	Advanced persistent threat
CNN	Convolution Neural Network
CSS	Cascading Style Sheets
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IA	Intelligence Artificial
IDS	Intrusion detection system
IOT	Internet of things
IP	Internet Protocol
LFI	Local File Inclusion
MITM	Man in the Middle
NLP	Natural Language Processing
NLU	Natural Language Understanding
OSI	Open Systems Interconnection
PDF	Portable Document Format
RNN	Recurrent Neural Network
SQL	Structured Query Language
UML	Unified Modeling Language
URL	Uniform Resource Locator
VPN	Virtual Private Network
XML	Extensible Markup Language

Introduction Générale

Vols, escroqueries et espionnages se sont développés sans commune mesure depuis l'apparition d'Internet. En effet, sur la toile, rien n'est, semble-t-il, plus facile pour les pirates de récupérer des numéros de cartes bancaires, des documents à caractère personnel ou encore des données stratégiques appartenant à des entreprises, selon un article paru sur Ouest-France, les cybers menaces se concentrent davantage sur les entreprises françaises que sur les particuliers, puisqu'elles récoltent 77 % des attaques. [30]

Aujourd'hui les systèmes informatiques se développent rapidement, deviennent de plus en plus un besoin incontournable pour les entreprises que maintenant ces derniers peuvent confier leurs données importantes et secrètes à ces techniques pour être manipulée ce qui a donner envie a plusieurs entreprise d'investir dans les formations de la sécurité informatique. Il est ainsi pertinent de s'intéresser à des idées plus fiables que les formations comme tests d'intrusions et les simulations dans la sécurité informatique.

Mais l'apprentissage avec un très grand nombre de ressources et supports et encore le manque de suivi ralenti le développement, pour cela un système qui répond à ces problèmes est trop demandée.

Dans ce mémoire nous allons s'intéresser sur la lutte contre les cyberattaques en intégrant l'apprentissage par la pratique dont on utilise le concept des tests d'intrusions et de la simulation sur les côtés défenses et attaques, ainsi qu'une assistance intelligente pour guider cet apprentissage.

Afin de traiter le sujet et d'intégrer un tel apprentissage une étude sur les utilisateurs dans le domaine de sécurité a été établi en prenant leur statistiques des différentes plateformes.

Ainsi à partir de notre système et les données générées par ce dernier concernant les utilisateurs nous voudrions voir l'avancement et les avis des utilisateurs concernant le domaine et les difficultés qui reculent l'apprentissage.

Nous verrons dans un premier temps une vision générale sur la sécurité informatique et ces différents domaines (Chapitre 1). Ensuite expliquer les idées des tests d'intrusion et des simulations et leur importance sur le domaine en montrant les avantages et les inconvénients de ces derniers (Chapitre 2). Avant de terminer une vue globale sur l'intelligence artificielle et comment une assistance intelligente peut être mis en place est expliquée (Chapitre3). Enfin une vision conceptuel et un processus de la plateforme est présentée (Chapitre 4).

Chapitre 1 : Généralité sur la sécurité informatique

1.1 Introduction

De nos jours l'internet est devenue un outil indispensable pour l'humain, un outil indispensable pour la recherche, l'analyse et le traitement de l'information et aussi pour la communication quel que soit les distances qui séparent les hommes.

Ce savoir et cette nouvelle technologie sont utilisés par des bon ou des mauvais utilisateurs, et avec l'évolution de la technologie de l'information, l'internet est devenu une source pleine de richesse et d'informations personnels et sensibles ce qui l'appétit des mauvaises personnes afin de récupérer les informations et les réutiliser pour des fins personnelles et surtout mal intentionnées.

Le développement rapide et la croissance des échanges entre les populations et la modalisation ont nécessité le développement des moyens de communication beaucoup plus fiables et beaucoup plus performants.

Les 'évolutions de la technologie qui fait du Net un moyen incontrôlable. Les entreprises ont ouvert leurs systèmes d'information à leurs partenaires et leurs fournisseurs ce qui a créé d'autres problèmes. Afin de corriger ces problèmes, les entreprises ont été obligées de trouver des solutions techniques pour maitriser, Controller et protéger tout type informations sensibles.

Pour parer à cette problématique il est important de connaitre et évaluer les différentes menaces, risques, vulnérabilité et surtout bien connaitre et maitriser les bases de la sécurité de l'information.

1.2 Définition de la sécurité informatique

La sécurité informatique est un processus continu de protections des objets et des données, elle consiste à empêcher tout accès ou action non autorisée aux ressources et garantir que les ressources sont utilisées par les objets et des personnes autorisés.

1.3 Objectifs et Critères fondamentaux de la sécurité

La sécurité informatique consiste d'une manière générale que les ressources matérielles et logicielles d'une organisation sont uniquement utilisées dans le cadre prévu.

Nous allons détailler les différents critères de la sécurité des données dans un système d'information.

Confidentialité

C’est-à-dire garantir que les données sont bien celles que l’on croit être.

Intégrité

Ensemble des mécanismes garantissant qu’une information n’a pas été modifiée. [1]

Disponibilité

La disponibilité est la caractéristique d’une information d’être accessible et utilisable par son destinataire autorisé à l’endroit et à l’heure prévue.

Traçabilité

Est la caractéristique qui conserve les traces de l’état et des mouvements de l’information. Sans elle, on n’a aucune chance d’avoir l’assurance que les trois autres critères sont respectés. [2]

Non répudiation

Mécanisme permettant de garantir qu'un message a bien été envoyé par un émetteur et reçu par un destinataire.

1.4 Politique de sécurité

On appel politique de sécurité un ensemble de règle qui fixent les actions autorisé et interdite dans le cadre d'un domaine de sécurité.

Elle fixe les principaux paramètres notamment les niveaux de tolérances et les couts acceptables, la fiabilité et les effets de la sécurité sur nos objets.

1.5 Mécanismes de sécurité

La sécurité des systèmes informatique s'intéresse aux mécanismes qui évitent les intrusions, dans cette section on va expliquer les différents mécanismes de sécurité utilisée aujourd'hui.

1.5.1 Système de détection d'intrusion

(*Intrusion Detection System*) un mécanisme écoutant le trafic réseau de manière furtive afin de repérer des activités anormales ou suspectes et permettant ainsi d'avoir une action de prévention sur les risques d'intrusion. [5]

1.5.2 Antivirus

Tout simplement un antivirus est un logiciel capable de détecter les virus informatiques et de les éliminer en utilisant certains algorithmes prédéfinis, une simulation des virus et des malwares peuvent être efficace pour l'amélioration d'un antivirus.

1.5.3 Firewall

Un firewall est un système physique (matériel) ou logique (logiciel) servant d'interface entre un ou plusieurs réseaux afin de contrôler et éventuellement bloquer la

circulation des paquets de données, en analysant les informations contenues dans les couches 3, 4 et 7 du modèle OSI.

Il s'agit donc d'une machine (machine spécifique dans le cas d'un firewall matériel ou d'un ordinateur sécurisé hébergeant une application particulière de firewall) comportant au minimum deux interfaces réseau :

• une interface pour le réseau à protéger (réseau interne).

• une interface pour le réseau externe.

1.5.4 Réseau virtuel privé (VPN)

Les VPN sont des tunnels de communication et ces tunnels peuvent être chiffrés ! Ils peuvent alors jouer le rôle du HTTPS et protéger vos communications des attaques MITM. C'est vrai, mais un VPN peut aussi déboucher sur l'ensemble d'Internet et donc vous permettre de naviguer sur le Web à travers lui. [6]

Un VPN peut servir à votre utilisateur pour entrer juste dans la topologie et le réseau de la simulation et systèmes virtuelle.

1.5.5 Cryptographie

La cryptographie est une des disciplines de la cryptologie s'attachant à protéger des messages (assurant confidentialité, authenticité et intégrité) en s'aidant souvent de secrets ou clés.

1.5.5.a Cryptographie symétrique

(Également dite a clé secrète), est la plus ancienne forme de Chiffrement. Elle permet à la fois de chiffrer et de déchiffrer des messages à l'aide d'une même clé.

1.5.5.b Cryptographie asymétrique

(Où cryptographie à clé publique), est une méthode de chiffrement qui s'oppose à la cryptographie symétrique. Elle repose sur l'utilisation d'une clé publique.

(Qui est diffusée) et d'une clé privée (gardée secrète), la 1ere permet de chiffrer le message et L’autre de le déchiffrer.

1.5.6 Signature numérique

La signature numérique est un mécanisme permettant de garantir l'intégrité d'un document électronique et d'en authentifier l'auteur, par analogie avec la signature manuscrite d'un document papier. Elle se différencie de la signature écrite par le fait qu'elle n'est pas visuelle, mais correspond à une suite de caractères. Elle ne doit pas être confondue avec la signature électronique manuscrite.

1.5.6.a Fonctions de la signature numérique [10]

Un mécanisme de signature numérique doit présenter les propriétés suivantes :

Il doit permettre au lecteur d'un document d'identifier la personne ou l'organisme qui a apposé sa signature (propriété d'identification).

Il doit garantir que le document n'a pas été altéré entre l'instant où l'auteur l'a signé et le moment où le lecteur le consulte (propriété d'intégrité).

Pour cela, les conditions suivantes doivent être réunies :

Authentique : l'identité du signataire doit pouvoir être retrouvée de manière certaine ;

Infalsifiable : la signature ne peut pas être falsifiée. Quelqu'un ne peut se faire passer pour un autre ;

Non réutilisable : la signature n'est pas réutilisable. Elle fait partie du document signé et ne peut être déplacée sur un autre document ;

Inaltérable : un document signé est inaltérable. Une fois qu'il est signé, on ne peut plus le modifier ;

Irrévocable : la personne qui a signé ne peut le nier.

En pratique, l'essentiel des procédures de signature numérique existantes s’appuie sur la cryptographie asymétrique, dans le reste de l'article nous nous placerons dans ce cas le plus courant. Les exemples d'échanges de données sont illustrés par les personnages Alice et Bob.

La figure ci-dessous illustre le fonctionnement de la signature numérique **(Voir Figure 1)**

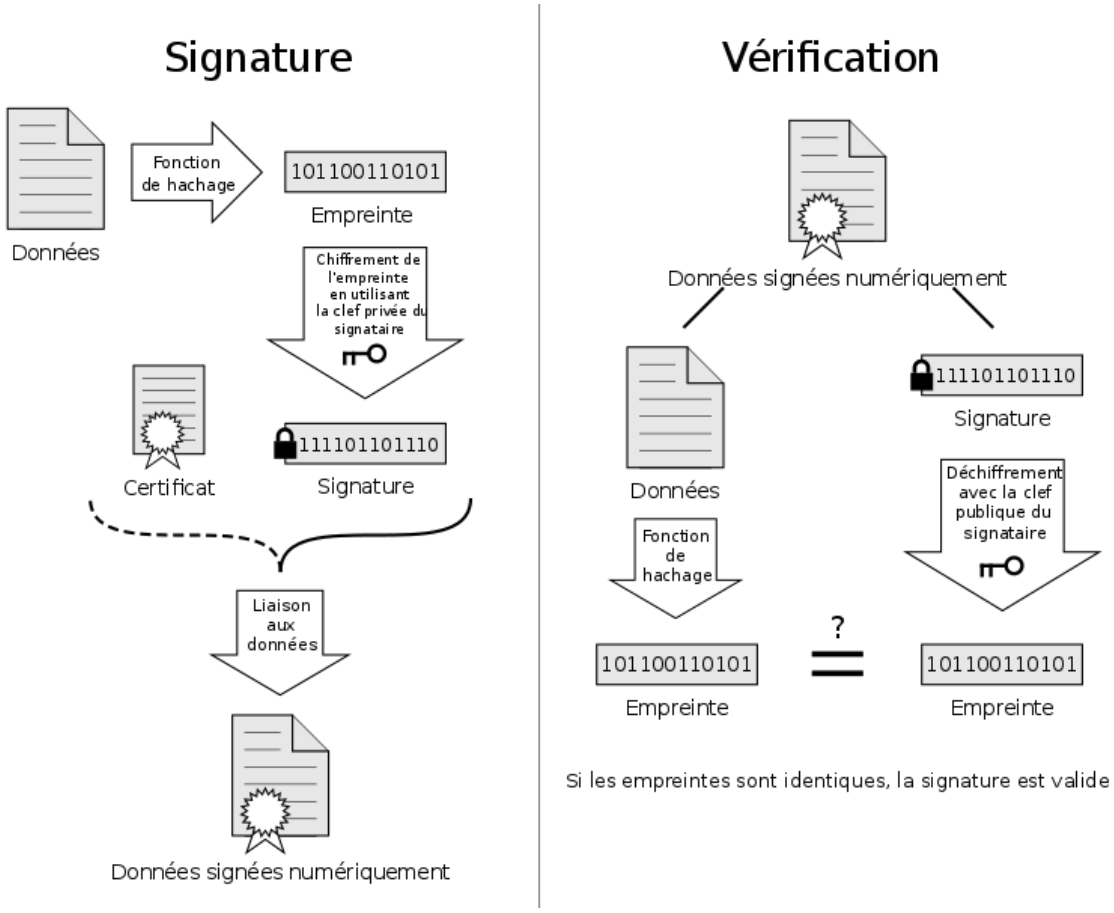

Figure 1 - Fonctionnement de la signature digitale [10]

Des variantes existent comme les K parmi N, où la signature est considérée comme valable si K membres du groupe parmi les N définis ont signé. Ce système sera utilisé par exemple lorsque l'autorisation de plusieurs services sera nécessaire pour déclencher un dispositif d'une gravité dépassant les prérogatives de chacun d'eux. Tel serait le cas par exemple pour une procédure de mise sur écoute téléphonique nécessitant les accords à la fois d'une instance autorisée de l'exécutif et d'une instance autorisée du législatif. On interdit ainsi l'usage de renseignements d'États à des fins personnelles, puisque le déblocage nécessite une coordination externe qui sera donc elle-même tracée.

L'intégration de la signature numérique sera vue et expliquée en détails dans le chapitre 4 (Conception et la mise en œuvre) dans la partie des certifications.

1.6 Menaces Informatique

1.6.1 Définition d'une attaque

Une attaque est une technique utilisée pour exploiter une ou plusieurs failles d'un système informatique à des fins non connues par l'exploitant du système et elle est généralement destructive.

Aujourd'hui sur le monde d'internet des attaques ont lieu chaque jour, à raison de plusieurs attaques par minute sur chaque machine connectée. Ces attaques sont pour la plupart lancées automatiquement à partir de machines infectées (par des virus, chevaux de Troie, vers, etc.) appelées pc zombie, à l'insu de leur propriétaire. Plus rarement il s'agit de l'action de pirates informatiques.

1.6.2 Types des attaques

1.6.2.A Attaques active [3]

Les attaques actives sont les attaques dans lesquelles l'attaquant tente de modifier l'information ou crée un faux message. La prévention de ces attaques est assez difficile, en temps réel, en raison d'un large éventail de vulnérabilités physiques, des réseaux et des logiciels. Au lieu d'une action préventive, l'administrateur met l'accent sur la détection de l'attaque et la récupération de toute perturbation ou retard causé par celui-ci. **(Voir Figure 2)**

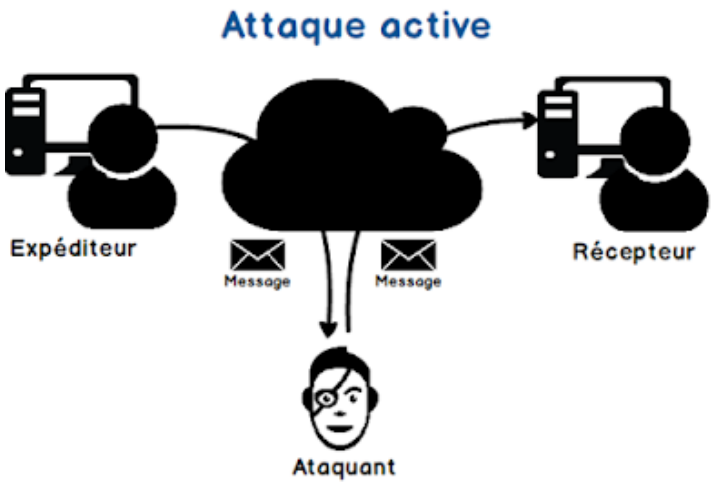

Figure 2 - Attaque active [3]

1.6.2.b Attaques passive [3]

Les attaques passives sont les attaques où l'attaquant se met en écoute non autorisée, en surveillant simplement la transmission ou la collecte d'informations. L'oreille indiscrète n'apporte aucun changement aux données ou au système. **(Voir Figure 3)**

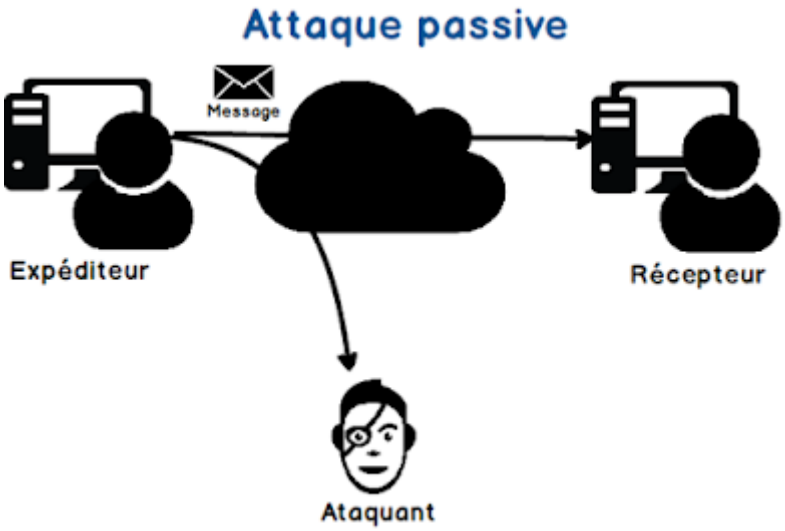

Figure 3 - Attaque passive [3]

1.6.3 Définition de vulnérabilité

La vulnérabilité est une faiblesse à la conception, la configuration ou la mise en œuvre d'un système.

1.6.4 Types des vulnérabilités

1.6.4.a Vulnérabilité des données [4]

La facilité d'espionnage et de trucage des communications résulte du fait que la plupart du trafic qui transite sur les réseaux, s'effectue « en clair », c'est-à-dire sans procédé de chiffrement. Par conséquent, les lignes de communication et donc les transferts de messages électroniques (e-mail), de mots de passe, de fichiers peuvent être surveillé et enregistrer à l'aide de logiciels spécialisés.

1.6.4.b Les vulnérabilités dues à l'absence d'une politique de sécurité [4]

Il n'est pas rare de constater qu'un réseau d'entreprise, par exemple, autorise plus de services entrée/sortie que nécessaire. Or, il est primordial de limiter l'accès à ces services qui peuvent permettre à un intrus connaissant bien le réseau interne d'obtenir des informations précieuses pour sa tâche d'espionnage ou de sabotage. Il est donc nécessaire d'établir une politique de sécurité définissant les restrictions d'accès et d'utilisation des services à appliquer.

1.6.4.c Les vulnérabilités liées aux erreurs de configuration [4]

Le paramétrage de dispositifs de sécurités telles que des routeurs filtres permettant grâce à des listes d'accès (Access List) de limiter l'accès à des services, est souvent complexe et peut entraîner des erreurs de configuration accidentelles. De telles erreurs peuvent réduire à néant l'efficacité d'une politique de sécurité.

1.7 Autres termes dans la sécurité informatique

1.7.1 Exploitation Web

Les sites et les applications web sont très utilisés de nos jours, ces derniers peuvent contenir des objets et des données très importants et sensibles, qui peuvent être militaire ou bancaire ou autres données secrètes, néanmoins l'environnement du web est souvent vulnérable et fragile à des attaques qui peuvent être exploitées sur le Net soit par des pirates informatique ou d'autres personnes malveillantes.

On trouve dans les exploitations des sites web de nombreux exemples des vulnérabilités :

Injection SQL : L'injection SQL est une méthode d'attaque très connue. C'est un vecteur d'attaque extrêmement puissant quand il est bien exploité. Il consiste à modifier une requête SQL en injectant des morceaux de code non filtrés, généralement par le biais d'un formulaire.

XSS (Cross Site Scripting) : L'attaque XSS repose sur ces problématiques. Elle est possible lorsqu'une valeur qui peut être contrôlée par l'utilisateur est injectée dans une page web sans suffisamment de contrôles, et que cette valeur peut être du code HTML/JavaScript valide, qui sera alors interprété par le navigateur.

LFI (Inclusion de fichier local) : La faille LFI tient son nom de Local File Inclusion (Inclusion de fichier local). Elle permet à un utilisateur d'inclure des fichiers locaux (appartenant donc au serveur externe) à partir d'une URL.

1.7.2 Cryptanalyse

Elle s'agit de l'étude des mécanismes théoriques et techniques visant à briser (casser) les algorithmes de chiffrements, c'est à dire il faut retrouver le message claire **M** à partir de message cryptée **C** sans connaitre la clé **K**.

On parlant d'attaques cryptanalyse, il existe 4 grands types :

Attaque sur texte chiffré seul : le cryptanalyste possède des exemplaires chiffrés des messages, il peut faire des hypothèses sur les messages originaux qu'il ne possède pas. La cryptanalyse est plus ardue de par le manque d'informations à disposition.

Attaque à texte clair connu : le cryptanalyste possède des messages ou des parties de messages en clair ainsi que les versions chiffrées. La cryptanalyse linéaire fait partie de cette catégorie.

Attaque à texte clair choisi : le cryptanalyste possède des messages en clair, il peut créer les versions chiffrées de ces messages avec l'algorithme que l'on peut dès lors considérer comme une boîte noire. La cryptanalyse différentielle est un exemple d'attaque à texte clair choisi.

Attaque à texte chiffré choisi : le cryptanalyste possède des messages chiffrés et demande la version en clair de certains de ces messages pour mener l'attaque.

1.7.3 Reverse Engineering [7]

Le Reverse Engineering, ou rétro-ingénierie / ingénierie inverse en français, représente l'étude et l'analyse d'un système pour en déduire son fonctionnement interne, et se retrouve dans de nombreux domaines de l'ingénierie : génie civil, mécanique, ingénierie navale, aéronautique.

Dans le domaine informatique, faire du reverse revient souvent à utiliser des outils d'analyse comme le désassembleur ou le décompileur dans le cadre d'analyse en boîte blanche. Nous reviendrons par la suite sur les différents contextes d'analyse. On retrouve également des outils permettant d'analyser les entrées/sorties d'un programme en sniffant le réseau comme Wireshark, ainsi que de nombreux autres outils, sur lesquels nous reviendront par la suite.

Le reverse étant très lié au monde de la sécurité informatique, une distribution Linux appelée Kali Linux (anciennement Backtrack) spécialisée dans le domaine de la sécurité est un très bon point de départ pour se former au monde du reverse. S'y aventurer n'est pas chose facile, mais de nombreux sites web, conférences (Nuit du Hack, Hack in Paris, DEFCON, etc...), livres, et autres supports vous aideront durant cette longue quête.

1.7.4 Exploitation Binaire

Les binaires, ou les exécutables, sont du code (instructions) machine à exécuter par un ordinateur. Pour la plupart, les binaires qu'on rencontre sont des fichiers ELF Linux ou des exécutables, sous Windows occasionnellement. L'exploitation du binaire est un vaste sujet au sein de cyber sécurité qui revient vraiment à trouver une vulnérabilité dans le programme et à l'exploiter pour prendre le contrôle d'un Shell ou modifier les fonctions du programme. [8]

Dans l'exploitation binaire on trouve les terminologies et les concepts suivants [8] :

Tampons (Buffers): Un tampon est un espace alloué dans la mémoire où les données (souvent saisies par l'utilisateur) peuvent être stockées, Elles peuvent êtres des variables globales ou dynamiquement allouées par le tas (Heap).

La pile (Stack) : Dans l'architecture informatique, une **pile** (en anglais stack) est une structure de données fondée sur le principe « dernier arrivé, premier sorti » (une file d'attente, FIFO), Dans x86, la pile est simplement une zone de la RAM qui a été choisie comme pile - il n'y a pas de matériel spécial pour stocker le contenu de la pile. Le registre esp / rsp contient l'adresse en mémoire où réside le bas de la pile. Quand quelque chose est poussé vers la pile, esp décrémente par 4 (ou 8 sur x86 64 bits), et la valeur qui a été poussée est stockée à cet emplacement en mémoire. De même, lorsqu'une instruction pop est exécutée, la valeur à esp est récupérée (c'est-à-dire que esp est déréférencé), et esp est ensuite incrémenté de 4 (ou 8).

Le tas (Heap) : Le tas est un endroit en mémoire. Un programme peut l'utiliser pour créer dynamiquement des objets. La création d'objets sur le tas présente certains avantages par rapport à l'utilisation de la pile.

1.7.5 Investigation numérique (Digital Forensics) [9]

Le digital forensics est une branche de la science qui se concentre sur la récupération et l'investigation de matériel trouvé dans des appareils numériques liés à la cybercriminalité. Le terme digital forensics a été utilisé pour la première fois comme synonyme de forensics informatique. Depuis lors, il s'est étendu pour couvrir l'investigation de tout appareil capable de stocker des données numériques. Bien que le premier crime informatique ait été signalé en 1978, suivi de la loi de Floride sur les ordinateurs, ce n'est que dans les années 1990 qu'il est devenu un terme reconnu. Ce n'est qu'au début du 21e siècle que des politiques nationales sur le digital forensics ont vu le jour.

Les Etapes du digital forensics :

Identification : Tout d'abord, trouvez les preuves, notez où elles sont stockées

Préservation : Ensuite, isolez, sécurisez et préservez les données. Cela comprend empêcher les gens de falsifier les preuves.

Analyse : Ensuite, reconstruisez des fragments de données et tirez des conclusions sur la base des preuves trouvées.

Documentation : Ensuite, créez un enregistrement de toutes les données pour recréer la scène du crime.

Présentation : Enfin, résumez et tirez une conclusion.

Toutes les catégories citées précédemment seront intégré dans notre plateforme, l'utilisateur va pouvoir apprendre tous ces domaines.

1.8 Conclusion

En conclusion La sécurité doit être un processus transverse à tout système présentant des risques et supportant des menaces, elle joue un rôle important et très particulier puisque la moindre défaillance peut compromettre le bon fonctionnement du système, pour cela dans ce chapitre on a élaboré les différents principes de sécurité informatique, les critères et comment peut-on atteindre les objectifs de la sécurité en utilisant les différents dispositifs tel l'antivirus les IDS.

Après avoir une vision générale sur la sécurité informatique et les différents mécanismes et concepts utilisés pour se protéger, d'autres méthodes existe pour optimiser et renforcer la sécurité des systèmes informatique, on parle donc des tests d'intrusion et des simulations des attaques, c'est deux concepts seront expliqués en détail dans le chapitre suivant (Chapitre 2).

Chapitre 2 : La Simulation et le test d'intrusion

2.1 Introduction

Le XXI [e] **siècle**, la cyber sécurité s'impose comme l'un des domaines le plus intéressant et l'un des secteurs les plus attractifs, elle est devenue rapidement un domaine très intéressant à apprendre par les informaticiens et aussi les autres personnes dans les domaines techniques surtout, mais la multitude des ressources et des documents sur internet perturbe les choix d'apprentissage et le rend plus complexe. Pour cela la majorité des apprentis suit une approche différente d'apprentissage qui s'appelle l'apprentissage par la pratique (Learning by doing en anglais).

L'apprentissage par la pratique quant à lui est une approche d'alternation entre la théorie et la pratique avec la mise en situation en milieu professionnel, c'est une méthode qui encourage la motivation chez les utilisateurs.

Dans ce chapitre nous allons expliquer deux nouveaux concepts de base dans l'approche de l'apprentissage par la pratique :

- La simulation des cyberattaques et les tests d'intrusion
- La simulation des attaques contre les défenses « Red Team vs Blue Team ».

2.2 Test d'intrusion

2.2.1 Qu'est-ce qu'un test d'intrusion ? [11]

Les tests d'intrusion ou « pentest » sont conçus pour tester la robustesse d'un système, un réseau ou une application sur tout type de plate-forme afin de détecter des écarts ou des vulnérabilités susceptibles d'être exploitées par un intrus ou un pirate informatique.

Les tests d'intrusion sont essentiels pour comprendre clairement les écarts d'une application ou d'un système en matière de sécurité, l'impact si ces écarts sont exploités lors d'une attaque et un plan clair et hiérarchisé basé sur les risques pour traiter les vulnérabilités rapidement.

Les tests d'intrusion peuvent impliquer des parties de l'environnement d'une organisation ou de tout son environnement. Ils peuvent tirer parti des pirates « chapeau » ou « chapeau noir » et peuvent être menées manuellement ou presque entièrement avec des outils automatisés.

2.2.2 Différence entre un audit de sécurité et un test d'intrusion

La différence entre un test d'intrusion et un audit est la motivation pour le consultant à analyser la machine cible et allez jusqu'à exploiter les failles de vulnérabilité et les montrer, il permet ainsi de situer le degré du risque de cette vulnérabilité.

2.2.3 Principe de base

Le principe consiste généralement a bien analyser et comprendre l'infrastructure du système ou de l'architecture réseau ciblée , afin d'imaginer ou de simuler l'attaque d'un pirate ou d'un utilisateur mal intentionné et avoir une idée générale des parties vulnérable qui peuvent donner un point aux pirates.

Le consultant « pentester » de cyber sécurité analyse le système au complet et détecte les risques potentiels dus à une mauvaise configuration d'un système d'information , d'un défaut d'une conception mal fait , de fautes de programmation ou encore des vulnérabilités liée a de manque d'une mise à jour.

Le consultant « pentester » donc lors d'un test d'intrusion, il se met dans la position d'un pirate. Le but de cette manœuvre est de trouver des vulnérabilités qui peuvent être exploitées pour proposer des actions qui peuvent corriger cette vulnérabilité et d'améliorer la sécurité du système pour le rendre plus puissant contre les attaques et empêcher les pirates informatique de pénétrer ou compromettre les infrastructures internes et externe des systèmes de l'entreprise.

Le test d'intrusion et l'analyse peut se réaliser selon trois cas, selon la demande des clients :

- le testeur ne possède aucune information et aucun code des applications du système.

- le testeur possède un nombre très limité d'informations comme les données d'authentifications etc...

- le testeur possède trop d'information comme le code source des applications, la base de données, l'architecture réseaux etc. ...

2.2.4 Types des tests d'intrusion

Comme vu précédemment dans le principe du test d'intrusion que l'utilisateur peut se mettre à l'une des trois positions qui sont appelée test de boîte noir, boîte blanche et boîte grise.

2.2.4.b Test de la boîte noire (Black Box)

Dans un test d'intrusion en mode boîte noir , le testeur n'a aucune information sur le système , il commence à chercher les informations de la version du système , les informations de l'entreprise et des employées , ensuite il commence à envoyer des requêtes ou des demande aux applications du système, suite à ses demandes il commence à analyser les réponses de ce dernier afin d'imaginer la composition globale du système et avoir une idée sur son comportement réel et notamment du code écrit derrière cela , afin de prévenir des éventuelles attaques et s'y protéger.

Exemples : une application vulnérable à la faille « SQL Injection » a l'authentification, donnée **sans le code source** ou se trouve les traitements de SQL et **sans les informations de connexion**.

2.2.4.b Test de la boîte grise (Gray Box)

En général, lors de tests d'intrusion en mode boîte grise « Grey box », le testeur dispose uniquement d'un couple identifiant - mot de passe. Ceci lui permet notamment de passer l'étape d'authentification. [11]

L'objectif de ce type de test est d'évaluer le niveau de sécurité vis-à-vis d'un « utilisateur normal ». [11]

Exemples : une application vulnérable à la faille « SQL Injection » à l'authentification, donnée sans le code source ou se trouve les traitements de SQL **mais les informations de connexion sont disponible**.

2.2.4.c Test de la boîte blanche (White Box)

Le testeur peut être au courant et ayant de nombreuses informations (mode « White box »). Parmi elles, les plus courantes sont : les schémas d'architecture, le compte utilisateur (permettant de s'authentifier), le code source de l'application, etc... [11]

Dans ce cas, il n'aura plus qu'une chose à faire : rechercher où sont les failles, et trouver le moyen de les exploiter. [11]

De même, un testeur se trouvant à l'intérieur du réseau à tester aura plus de facilité à trouver ces failles car il connaît non seulement le système, mais il peut avoir accès directement aux ressources dont il a besoin. [11]

Exemples : une application vulnérable à la faille « SQL Injection » donnée **avec le code source** ou se trouve les traitements de SQL.

Dans notre plateforme, nous allons développer un environnement d'apprentissage sur les trois types des tests d'intrusion pour donner une vue globale aux utilisateurs.

2.2.5 Les plateforme de test d'intrusion existante

Actuellement et suite à l'importante des tests d'intrusion dans l'informatique, de nombreuses plates-formes ont été mise en place pour la prise en charge des simulations des systèmes et des applications vulnérables consacrés aux utilisateurs afin d'améliorer les connaissances par la suite les compétences de l'utilisateur dans les tests d'intrusion.

2.2.5.a Qu'est-ce qu'une plateforme de test d'intrusion ?

Une plateforme de test d'intrusion c'est une plateforme qui propose aux utilisateurs des applications ou des systèmes virtuels avec des vulnérabilités pour que les utilisateurs puissent apprendre à faire des tests d'intrusion et avoir une idée générale sur le concept, on peut citer deux exemples de ce type de plateforme qui sont en premier lieu dans le marché de ce concept.

2.2.5.B La plate-forme root-me

La plateforme root-me est une plateforme rapide, accessible et réaliste pour tester les compétences dans le hacking, c'est une plateforme française crée en 2010, elle propose des différents challenges liés à la sécurité informatique dans les différentes catégories (Web, forensics, cryptographie, etc. ...) pour faciliter à l'utilisateur de comprendre le concept de la sécurité informatique

La plateforme donne un comportement simulé des applications dans les domaines cités précédemment, chaque application contient une vulnérabilité selon le titre du challenge

(Voir Figure 4), exploiter la vulnérabilité donne un texte qui s'appelle un drapeau « flag » en validant ce dernier **(Voir Figure 5)**, l'utilisateur reçoit des points selon la difficulté du challenge **(Voir Figure 6)**, ce concept est très intelligent pour motiver l'utilisateur à chercher de gagner toujours des points.

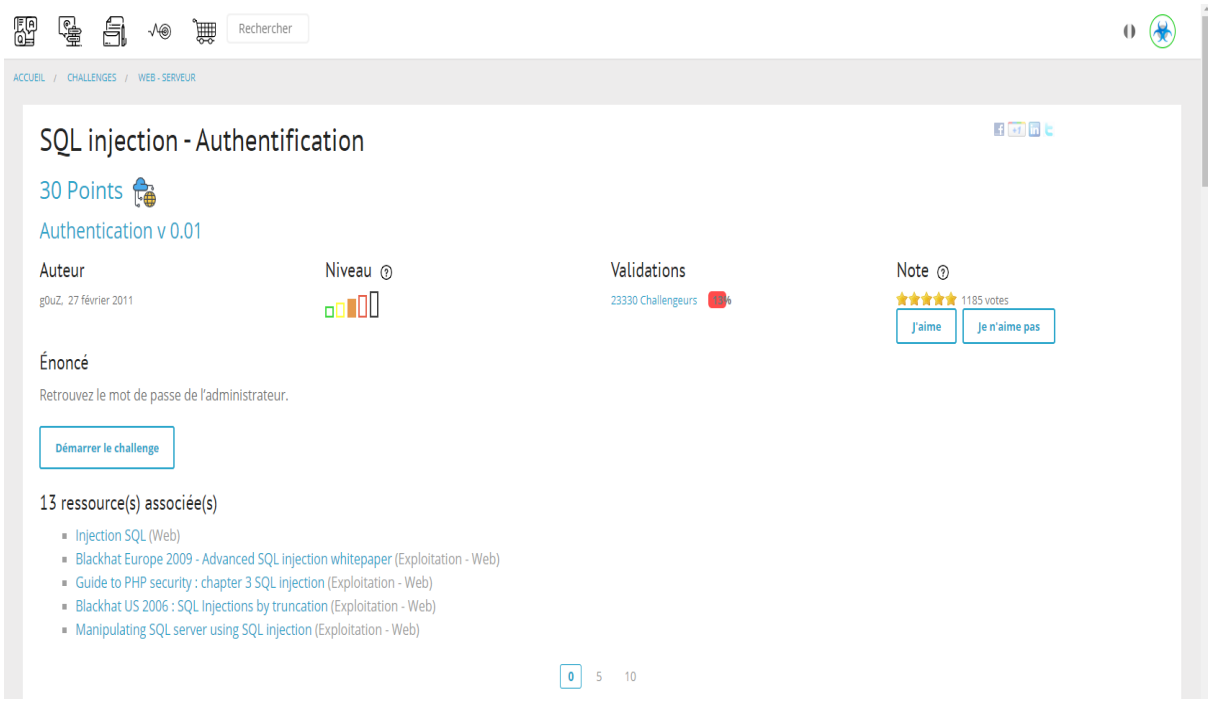

Figure 4 - Challenge dans la plateforme root-me

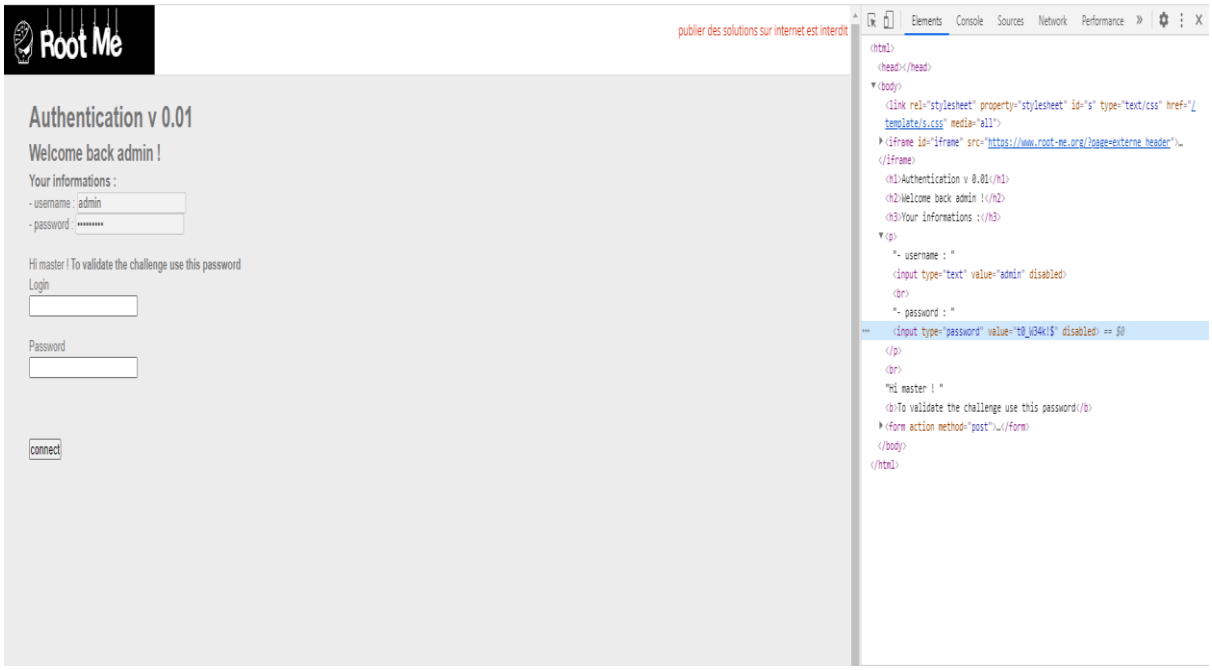

Figure 5 - Flag dans un challenge root-me

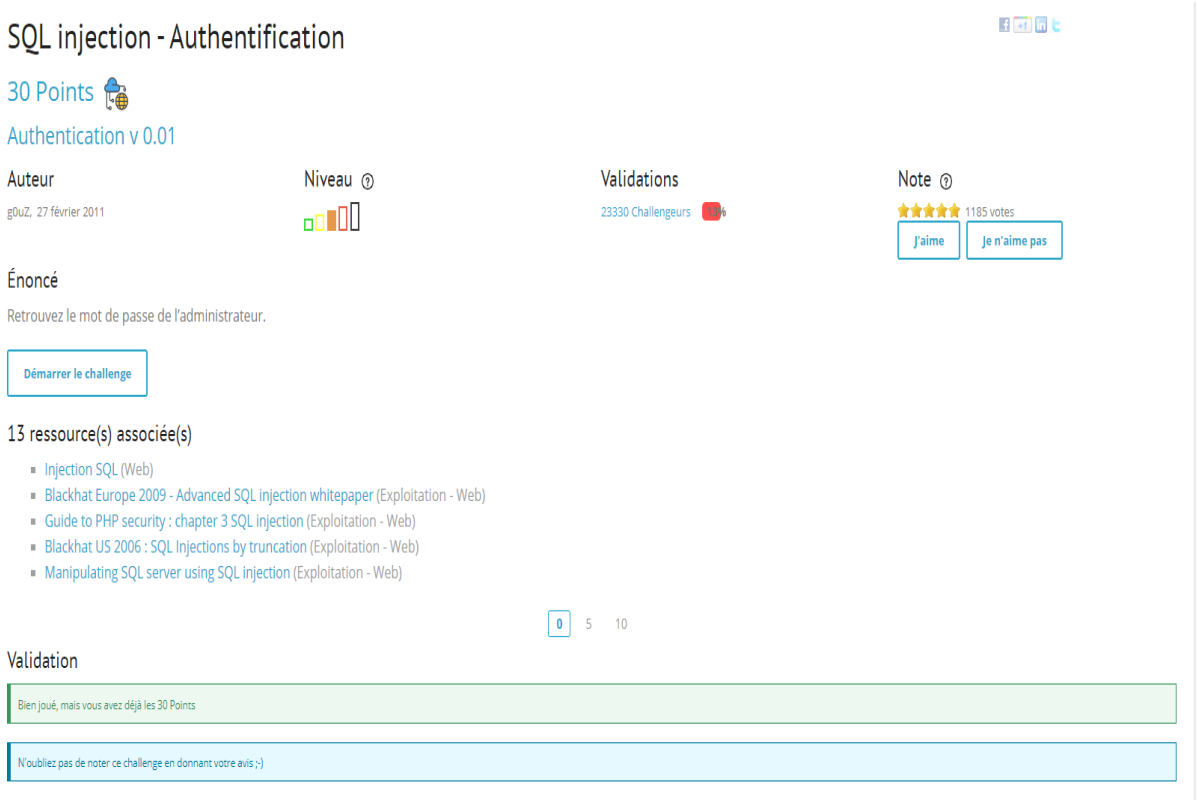

Figure 6 - Validation d'un flag dans root-me

2.2.5.c La plate-forme Hackthebox

Hack The Box est une plateforme en ligne dédiée à la sécurité informatique, elle permet de tester vos compétences en test d'intrusion et d'échanger des informations et des savoirs faire entre des milliers de personnes dans le domaine de la sécurité. L'inscription à la plateforme est précédé par le défi d'invitation, ou il faut pirater le code d'invitation dans la plateforme pour accéder, puis l'utilisateur peut se lancer dans l'une des nombreuses machines ou des défis en direct.

La plateforme propose une simulation des systèmes vulnérables , qui peuvent comporter une infrastructure complète avec une architecture réseaux bien définis , afin de mettre l'utilisateur dans une configuration du monde réel des tests d'intrusion des machines , l'utilisateur peut alors lancer l'une des machines proposée , alors il commence à analyser le système , jusqu'à obtenir deux drapeau , le drapeau de l'utilisateur normal d'un serveur , et le drapeau de l'administrateur d'un serveur (ou root dans un système linux) . **(Voir Figure 7)**

```
PS C:\Users\Administrator\Desktop> cmd /c dir /Q
cmd /c dir /Q
 Volume in drive C has no label.
 Volume Serial Number is 4638-2C29

 Directory of C:\Users\Administrator\Desktop

04/14/2019  11:35 AM    <DIR>          BUILTIN\Administrators .
04/14/2019  11:35 AM    <DIR>          NT AUTHORITY\SYSTEM    ..
03/27/2019  05:37 AM                34 RE\coby                root.txt
               1 File(s)             34 bytes
               2 Dir(s)  17,583,677,440 bytes free
```

Figure 7 - Exemple d'un challenge Hackthebox (validation d'un flag root)

2.2.6 Synthèse

Dans les sections précédentes nous avons vu les exemples des plateformes dans le marché , chaque plateforme a un très bon concept pour l'apprentissage des test d'intrusions , notre plateforme va intégrer les deux concepts cités dans les deux plateformes avec une possibilité de certifications dans certaines parties certifiée avec une signature numérique pour encourager et motiver les utilisateurs a mieux se développer et avancer sur les concepts.

Mise à part partie le concept des tests d'intrusion soit sur les challenges des applications, ou soit sure les systèmes complets vulnérables, la plateforme met à la possession des utilisateurs une partie des simulations des attaques et des défenses « Red Team VS Bleu Team ».

La simulation et les concepts de la simulation vont être expliqués dans la section suivante.

2.3 Simulation des cyber-attaques

2.3.1 Qu'est-ce qu'une simulation informatique ? [12]

Le mot simulation tient son origine du latin (Simulare), c'est faire semblant. Elle consiste également à reproduire le comportement du modèle dans son milieu et avec ses différents comportements et tout cela par le biais de la modélisation mathématique.

La simulation informatique ou numérique désigne l'exécution d'un programme informatique sur un ordinateur ou réseau en vue de simuler un phénomène physique réel et complexe (par exemple : chute d'un corps sur un support mou, résistance d'une plateforme pétrolière à la houle, fatigue d'un matériau sous sollicitation vibratoire, usure d'un roulement à billes…). Les simulations numériques scientifiques reposent sur la mise en œuvre de modèles théoriques utilisant souvent la technique des éléments finis. Elles sont donc une adaptation aux moyens numériques de la modélisation mathématique, et servent à étudier le fonctionnement et les propriétés d'un système modélisé ainsi qu'à en prédire son évolution. On parle également de calcul numérique. Les interfaces graphiques permettent la visualisation des résultats des calculs par des images de synthèse.

Ces simulations informatiques sont rapidement devenues incontournables pour la modélisation des systèmes naturels en physique, chimie et biologie, mais également des systèmes humains en économie et en science sociale, architecture. Elles permettent de limiter le risque et d'éviter le coût d'une série d'épreuves réelles (ex: essais de véhicules). Elles peuvent offrir un aperçu sur le développement d'un système trop complexe pour simuler avec de simples formules mathématiques (ex: ouragan).

2.3.2 Intérêts des simulations dans le domaine de la cybersécurité

La simulation dans le domaine dans la cyber sécurité a une grande importance pour comprendre le mécanisme des attaques, participé à une simulation sur une infrastructure réelle ou encore apprendre les pratiques essentielles de cyber défense.

Donc la simulation dans ce domaine a une grande part dans la sécurité des systèmes et des applications qui peuvent contenir des informations secrets et importantes.

2.3.3 Simulation « Red Team VS Blue Team»

La simulation « Red Team VS Blue Team » comme son nom l'indique elle oppose deux équipe , une rouge qui joue le rôle d'un attaquant et c'est une entité externe de l'entreprise , elle essaye de trouver et détecter les exploits pour casser la protection des systèmes de l'entreprise et l'autre bleu qui est une entité interne , elle essaye de protéger le système et de fermer toute les vulnérabilité et les parties faibles des systèmes et empêcher tout accès non autorisé . Plusieurs détails seront expliqués dans les sous sections suivante de ce document.

Cette simulation est très utile et importante dans le domaine de la sécurité, elle permet d'entrainer les employées sur les deux côtés, l'intégration de cette fonctionnalité et cette simulation sera vu et expliquée en détail dans le chapitre 4 (Conception et mise en œuvre).

2.3.4.a Red Team

« Red Team » ou l'équipe rouge en français, est une entité externe de l'entreprise, elle a le but d'évaluer la sécurité de l'entreprise, ou on peut dire tester le niveau de sécurité dans l'entreprise, leur travail et de chercher et détecter les exploits et les vulnérabilités dans les systèmes qui peuvent ouvrir des portes aux pirates.

La prestation se déroule sur plusieurs semaines. Cette période large permet d'intervenir par phase d'actions avec des attaques cyber ou encore des campagnes du « Phishing », en passant par l'intrusion physique des locaux par exemple. [15]

Les campagnes Red Team proposent d'évaluer d'une part le niveau de sécurité d'un système d'information de manière générale. D'autre part, elles permettent d'évaluer les actions de la Blue Team (l'équipe technique de l'entreprise ou de son SOC) face à la détection d'intrusions, quelle qu'elle soit. Les attaques doivent être complexes et préparées de manière à ne pas être repérées par l'équipe interne à l'entreprise. [15]

Quels sont les avantages pour une entreprise ? [15]

Grâce à cette approche, l'entreprise identifie un maximum de vulnérabilités susceptibles d'être exploitées, ainsi que les scénarios probables menant à une compromission du système d'information.

Les missions Red Team permettent également de sensibiliser les collaborateurs aux risques cyber. En ce sens, elles sont un axe de travail prioritaire pour l'amélioration du niveau de sécurité de l'entreprise.

2.3.4.b Blue Team

« Blue Team » ou l'équipe bleu en français, est une entité interne de l'entreprise, son rôle principale est d'aider l'entreprise à empêcher tout accès non autorisée, elle est similaire à la « Red Team » dans le sens où elle identifie les vulnérabilités possibles, la différence se place dans sa stratégie d'amélioration des mécanismes de défense. De plus, contrairement à la « Red Team », elle est au courant des défenses déjà en place. Elle est continuellement impliquée dans l'analyse d'activité suspicieuse.

Quels sont les avantages pour une entreprise ?

Grâce à cette approche, l'entreprise peut sécuriser les vulnérabilités de ses systèmes identifiées et construire des mécanismes de sécurité pour faire face aux attaques externes et internes.

Le principal avantage est l'amélioration continue de la posture de sécurité de l'organisation en identifiant les lacunes et en les comblant par des contrôles appropriés.

Concurrence entre les deux teams

En fonction de l'objectif de votre test, « la Red Team informera ou non « la Blue Team » d'un test prévu ou d'une action effectuée.

Par exemple, si l'objectif est de simuler un scénario de réponse réel à une menace « légitime », vous ne voudrez pas informer la Blue Team du test. Quelqu'un de la direction doit être au courant du test, généralement le chef de la Blue Team. Cela garantit que le scénario de réaction est toujours testé, mais avec un contrôle plus strict quand la situation est aggravée.

Lorsque le test est terminé, les 2 équipes recueillent des informations et font un rapport sur leurs conclusions.

La Red Team conseille la Blue Team si elle parvient à pénétrer les défenses, et donne des conseils sur la manière de bloquer des tentatives similaires dans un scénario réel.

De même, la Blue Team doit faire savoir à la Red Team si ses procédures de surveillance ont détecté ou non la tentative d'attaque.

Cette concurrence entre les deux teams crée une grande motivation et même une compétition entre eux pour que chaque un montre ces compétences.

2.3.4 Exemples des simulations des attaques

Vu l'importance des simulations des attaques dans le monde informatique, de nombreuse entreprises, et mêmes les organisations gouvernementales ont opté pour cette approche pour renforcer la sécurité de leur services. Certains exemples vont être cités dans les sous sections suivante pour montrer l'importance de ce concept.

2.3.4.a Simulation de cyber-attaques, Européens et Américains collaborent [13]

Près d'une centaine d'experts en informatique de 16 pays, européens et américains ont travaillé ensemble, dans le cadre d'un exercice de simulation, afin de contrer des cyber-attaques d'envergure ciblant notamment des agences de sécurité et des centrales électriques de l'Union européenne.

L'opération Cyber Atlantique 2011, est le premier exercice conjoint mené par l'UE et les États-Unis en matière de cyber sécurité. Deux scénarios ont été envisagés. Le premier consistait en une attaque ciblée et furtive de type APT (Advanced persistent threat) visant à soutirer des informations secrètes aux agences de cyber sécurité des États membres de l'Union et à les diffuser. Les experts en sécurité de l'ENISA, l'agence européenne chargée de la sécurité des réseaux et de l'information, ont déclaré que ce type d'attaques pouvait tout à fait se produire dans la réalité. « Ce genre de menace est typique de ce qui pourrait arriver, même si l'exercice n'est pas fondé sur une situation particulière. Nous avons choisi des menaces qui nous paraissaient réelles. Et nous n'avons pas choisi des attaques faciles à repousser », a expliqué Graeme Cooper, le porte-parole de l'ENISA.

Attaques sur les productions énergétiques

La seconde simulation concernait la perturbation des systèmes de contrôle et d'acquisition de données (SCADA) dans les infrastructures de production d'électricité. Cette menace est prise très au sérieux par les autorités de l'UE, d'une part à cause d'allégations selon lesquelles le groupe de pirates Anonymous aurait tenté d'infiltrer les centrales françaises et suite également à l'attaque par le ver Stuxnet qui a endommagé les installations nucléaires iraniennes. Plus de 20 pays de l'UE ont participé à l'exercice, dont 16 activement. La Commission européenne a dirigé l'opération et le département américain de la sécurité Intérieure a apporté son soutien aux équipes de sécurité mobilisées. L'un des objectifs était de voir comment l'UE et les États-Unis pouvaient se soutenir et coopérer en cas de cyber-attaques sur leurs infrastructures d'information critique. L'opération faisait également suite au premier stress test de sécurité informatique réalisé au niveau de l'Europe l'an dernier, Cyber Europe 2010.

2.3.4.b Simulation "Red Team VS Blue Team" du ministère des Armées en matière de cyberdéfense

A l'occasion de la cinquième conférence cyber sécurité de L'Usine Digitale, qui s'est tenue à Paris mardi 25 juin 2019, le commandement de la cyberdéfense (Comcyber) du ministère des Armées a détaillé sa stratégie de lutte contre les pirates informatiques. Exercices de simulation et recherche de failles : il s'agit d'une action quotidienne en faveur de la protection des données confidentielles et de la souveraineté nationale.

Quand on parle du cyber sécurité, on pense nécessairement données « top secret » et souveraineté nationale. Le ministère des Armées est pleinement engagé dans cette guerre du XXIe siècle. A l'occasion de la cinquième conférence cyber sécurité de L'Usine Digitale, qui s'est tenue à Paris le 25 juin 2019, son commandement de la cyberdéfense a présenté sa stratégie de lutte contre les pirates informatiques. "Il s'agit pour nous de défendre l'intégrité des systèmes et, par là-même, l'action militaire", souligne « Sébastien Bombal », conseiller de l'officier général du commandement de la cyberdéfense (Comcyber).

Anticiper les Scénarios d'attaques

Le ministère dispose de réseaux avec plus de 200 000 ordinateurs, dont il faut assurer la protection. « Début 2019, opération Bug Bounty a été mise en place », indique Sébastien Bombal. Bug Bounty ? C'est un programme – déjà employé par de grands groupes, tels que les GAFAM – qui permet à des "hackers éthiques" de fouiller le système informatique d'une organisation pour en dénicher les failles.

La prestation se déroule sur plusieurs semaines. Cette période large permet d'intervenir par phase d'actions avec des attaques cyber ou encore des campagnes de Phishing, en passant par l'intrusion physique des locaux par exemple.

Les campagnes Red Team proposent d'évaluer d'une part le niveau de sécurité d'un système d'information de manière générale. D'autre part, elles permettent d'évaluer les actions de la Blue Team (l'équipe technique de l'entreprise ou de son SOC) face à la détection d'intrusions, quelle qu'elle soit. Les attaques doivent être complexes et préparées de manière à ne pas être repérées par l'équipe interne à l'entreprise.

2.4 Déploiement des simulations et des tests d'intrusions

Dans cette section on va expliquer les principales bases du déploiement d'un système à test d'intrusion ou des simulations, plusieurs principes et techniques sont utilisés pour construire cela.

Déployer un système de simulation des attaques ou un système des tests d'intrusion peut se faire en deux modes, un mode en ligne qui utilise le cloud pour héberger ses systèmes, ou le mode hors ligne ou interne de l'entreprise, ce mode utilise le concept des machines virtuelles.

Dans cette section les bases citées vont être expliquée brièvement pour que le lecteur puisse comprendre les principaux concepts du déploiement.

2.4.1 Mode En Ligne

Dans ce mode tous les utilisateurs sur internet peuvent utiliser les simulations de l'entreprise soit employée ou autres, ce concept utilise le cloud pour héberger ces systèmes en utilisant des architectures réseaux bien définis et sécurisés.

Dans cette section on va expliquer les idées principales du Cloud vu qu'il est l'idée de base sur le déploiement et hébergement des systèmes sur internet en mode en ligne.

2.4.1.a Service cloud

Le terme « services cloud » désigne un large éventail de services fournis à la demande sur Internet aux entreprises et aux clients. Ces services sont conçus pour fournir un accès simple et accessible aux applications et aux ressources, sans qu'une infrastructure interne ou du matériel ne soit nécessaire. Que ce soit pour consulter leurs e-mails ou collaborer sur des documents, la plupart des employés utilisent les services cloud dans leur travail, consciemment ou non. [16]

2.4.1.b Comment les services cloud fonctionnent-ils ?

Les services cloud sont entièrement gérés par des prestataires de services et des fournisseurs de cloud computing. Ils sont mis à la disposition des clients à partir des serveurs des fournisseurs : l'entreprise n'a donc pas besoin d'héberger les applications sur ses propres serveurs sur site. [16]

2.4.2 Mode Hors ligne

Dans ce mode, il y a que les utilisateurs des entreprises ou les employées qui peuvent bénéficier de ces services, ce mode utilise les machines virtuelles locale pour héberger ces systèmes.

2.4.2.a Machines Virtuelles

Une machine virtuelle est un fichier informatique, généralement appelé image, qui se comporte comme un ordinateur réel. En d'autres termes, il s'agit d'un ordinateur créé à l'intérieur d'un ordinateur. Ce fichier s'exécute dans une fenêtre, comme tout autre programme, en offrant à l'utilisateur final une expérience identique à celle qu'il aurait sur le système d'exploitation hôte. La machine virtuelle est placée dans un « bac à sable » qui l'isole du reste du système, de sorte que les logiciels installés sur la machine virtuelle ne peuvent ni s'échapper, ni modifier l'ordinateur hôte. Cela produit un environnement idéal pour tester d'autres systèmes d'exploitation.

Il est possible d'exécuter plusieurs machines virtuelles simultanément sur un même ordinateur physique. Pour les serveurs, les divers systèmes d'exploitation fonctionnent côte à côte, avec un composant logiciel appelé hyperviseur pour les gérer, alors que les ordinateurs de bureau classiques n'utilisent qu'un seul système d'exploitation pour exécuter d'autres systèmes d'exploitation dans des fenêtres de programme qui leur sont propres. Chaque machine virtuelle fournit son propre matériel virtuel, à savoir les processeurs, la mémoire, les disques durs, les interfaces réseau et les autres périphériques nécessaires. Le matériel virtuel est ensuite mappé au matériel réel sur la machine physique, ce qui permet de réaliser des économies en réduisant les besoins en matériel, ainsi que les coûts de maintenance, d'alimentation et de refroidissement associés.[17]

Ce concept est très important pour l'intégration des systèmes vulnérables pour l'exploiter par les utilisateurs sans prendre des risques et sans faire des dégâts.

2.5 Conclusion

Les compétences dans le domaine de la cyber sécurité ne s'obtiennent pas seulement en regardant des vidéos sur ce dernier ou bien en lisant le documents ou même en apprenant par cœur, mais bien en appliquant immédiatement ce que l'on apprend à l'aide des simulations des cyberattaques et les tests d'intrusions , une motivation pour les utilisateurs à atteindre un but précis des objectifs des simulations et des tests d'intrusions et une demi entrée au mondes de la cyber sécurité réel amusante pour les utilisateurs.

Chapitre 3 : Les systèmes d'assistance intelligente

3.1 Introduction

Nous vivons dans un monde hyper connecté dans lequel chaque interaction, Allant d'un l'appel téléphonique à l'achat en passant par l'affichage d'une page web, s'ajoute à cela un océan illimité de données. Avec l'arrivée de l'internet des objets (IoT), les voitures, les réfrigérateurs, les vêtements et toute autres objets, peuvent être commandés avec une adresse IP et peuvent également générer des quantités de data supplémentaires chaque minute ou chaque seconde, exploiter ces données d'une manière intelligente peuvent donner un grand avantage dans le domaine de la sécurité informatique et suivre ou aider l'utilisateur d'une manière intelligente.

Toutes ces données peuvent être utilisées pour des fins commerciales, ajuster vos actions marketing et fournir le service personnalisée et immédiat attendu par vos clients et le suivi nécessaire pour donner une bonne expérience aux utilisateurs. Mais comment les entreprises peuvent transformer ces données massives sans fond en flux régulier d'informations pertinentes pour suivre et connaitre la difficulté des utilisateurs afin de répondre à leurs attentes ?

La réponse à cette question réside dans l'utilisation de l'intelligence artificielle (IA).

Dans ce chapitre nous allons aborder les différents concepts de base de l'intelligence artificielle, Machine Learning et le Deep Learning, ainsi que les principes des chatbots qui peuvent aider l'utilisateur a bien se localiser au niveau de la plateforme, et les systèmes de recommandations intelligents pour proposer aux utilisateurs des formations avec un suivi pour lui permettre d'atteindre son objectif principal, tout cela à partir des interactions des utilisateurs et leur statistiques données dans la plateforme.

3.2 Intelligence Artificielle

3.2.1 Définition

L'intelligence artificielle est un ensemble de théories et de techniques développant des programmes informatiques complexes, ayant la capacité d'imiter l'intelligence humaine comme savoir raisonner et apprendre de manière automatique, le but ultime de l'intelligence artificielle n'est pas de remplacer l'humain mais bien de le décharger afin qu'il puisse se concentrer sur des taches de plus en plus créatives ou agréables.

3.2.2 L'intelligence artificielle et la cybersécurité

L'intelligence artificielle est de plus en plus intégrée dans différents modules de sécurité réseau. L'algorithme peut être entraîné pour effectuer certaines actions prédéfinies lorsqu'une attaque se produit. Les experts mettent en garde, mais les programmes d'IA qui reposent sur de mauvais algorithmes peuvent causer de graves problèmes.

L'augmentation des attaques informatiques rend la tâche très difficile pour les équipes de sécurité. Même avec la diversité et la complexité des technologies malveillantes, l'identification et l'évaluation des cybers menaces nécessitent l'examen de grandes quantités de données et la recherche de signaux et de comportements suspects. Donc et comme expliqué dans notre introduction de ce chapitre, un moyen d'exploiter et examiner toute ses données massives et rendre l'information pertinente est l'utilisation de l'intelligence artificielle dans le domaine. Plusieurs usages de l'intelligence artificielle on était mise en place on cite comme exemples l'identification des menaces réseau, la surveillance des emails et l'intelligence artificielle pour combattre l'intelligence artificielle.

Dans notre cas on ne s'intéresse pas sur la détection des attaques ou identifier les cybers menaces mais on s'intéresse surtout sur un concept d'intelligence artificielle qui nous aide à suivre l'utilisateur, le guider et l'assister dans le choix de ces ressources et ses formations d'apprentissage dans le domaine de la cyber sécurité, de nombreux concepts vous seront expliqué tout au long de ce chapitre.

Actuellement deux autres approches sont très connues aussi dans le domaine, On a le Machine Learning qui utilisée partout dans les entreprises et les grandes compagnies, et également on a le Deep Learning qui est quant à lui plus répondu dans la recherche.

Les deux approches citées seront présenté dans les sections suivantes pour bien expliquer les concepts de base utilisés dans la réalisation de nos systèmes intelligents.

3.3 Le Machine Learning ou Apprentissage Automatique

3.3.1 Définition

Le Machine Learning consiste donc à faire en sorte que les machines que l'on crée soient capables d'apprendre à partir de données récoltées, d'analyser et d'apprendre des actions qu'elles font pour ainsi pouvoir prendre des décisions stratégiques de manière « autonome » sans que personne ne leur dise explicitement ce qu'il faut faire. [18]

Les méthodes et techniques d'apprentissage automatique sont divisées en plusieurs catégories, chacune utilisant un type d'apprentissage. L'explication de ces types d'apprentissage est dans la sous-section suivante.

3.3.2 Les types d'apprentissage

3.3.2.a L'apprentissage Supervisé

La majorité des apprentissages automatiques utilisent un apprentissage supervisé (Supervised Learning).

L'apprentissage supervisé consiste en des variables d'entrée (x) et une variable de sortie (Y). On utilise un algorithme pour apprendre la fonction de mapping de l'entrée à la sortie.

$$\mathbf{Y = f\,(X)}$$

Le but est d'appréhender si bien la fonction de mapping que lorsque vous avez de nouvelles données d'entrée (x), vous pouvez prédire les variables de sortie (Y) pour ces données.

C'est ce qu'on appelle l'apprentissage supervisé, car le processus d'un algorithme tiré de l'ensemble de données de formation (training set) peut être considéré comme un enseignant supervisant le processus d'apprentissage. [23]

3.3.2.b 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.

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.

L'apprentissage non supervisé comprend deux catégories d'algorithmes: Algorithmes de regroupement et d'association. [23] **(Voir Figure 8)**

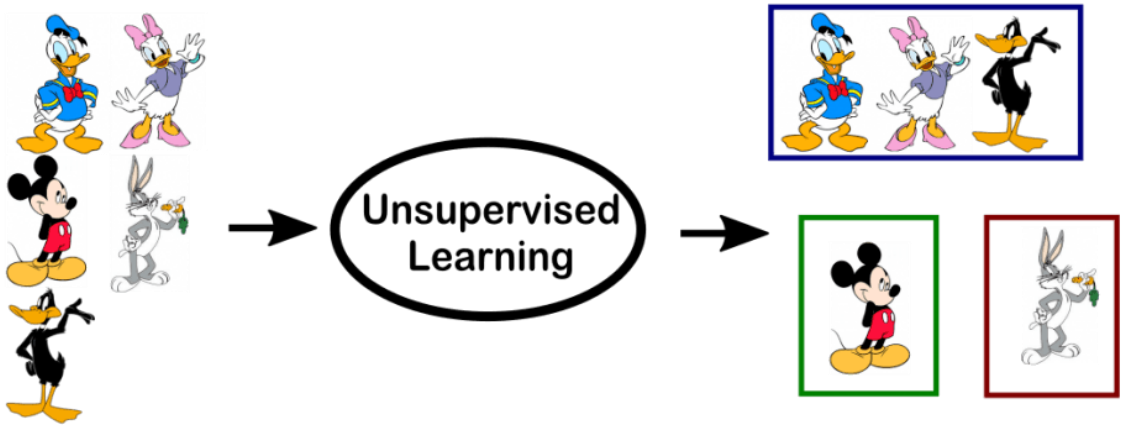

Figure 8 - Apprentissage non supervisé [23]

3.3.2.c Apprentissage semi-supervisé

Les problèmes qui consiste à avoir une grande quantité de données d'entrée (X) et que seules certaines données sont étiquetées (Y) sont appelés problèmes d'apprentissage semi-supervisés. Par conséquent, ces problèmes se situent entre l'apprentissage supervisé et l'apprentissage non supervisé. [23]

Exemple : une archive de photos dans laquelle seules certaines images sont étiquetées (chien, chat, personne, par exemple) et la plupart ne le sont pas.

3.3.2.d Apprentissage par renforcement

En intelligence artificielle, et plus précisément en Machine Learning, pour les agents autonomes (robots, etc.), l'apprentissage par renforcement comprend des actions d'apprentissage basées sur l'expérience pour optimiser les récompenses quantitatives au cours du temps.

3.3.3 Fonctionnement du Machine learning

Etape 1 : l'entraînement ou l'apprentissage

La première étape de la Machine Learning est d'entraîner le robot. Pour cela, il va falloir lui donner des exemples, de nombreux exemples. Le robot pourra ainsi observer et apprendre. L'idée est que la machine récupère de très nombreuses informations pour apprendre, les réutiliser pour s'adapter à de nouvelles situations voire même les anticiper avec l'enseignement qu'elle aura tiré des nombreux exemples donnés, d'où le terme d' « apprentissage ». [18]

Etape 2 : la réalisation de la mission

Le robot pourra alors construire, de manière autonome, ce qu'on appelle une « représentation interne », une sorte de cartographie intelligente de la situation pour pouvoir effectuer la tâche demandée et ainsi faire de la « prédiction », des « recommandations », des « prises de décisions », etc. [18] **(Voir Figure 9)**

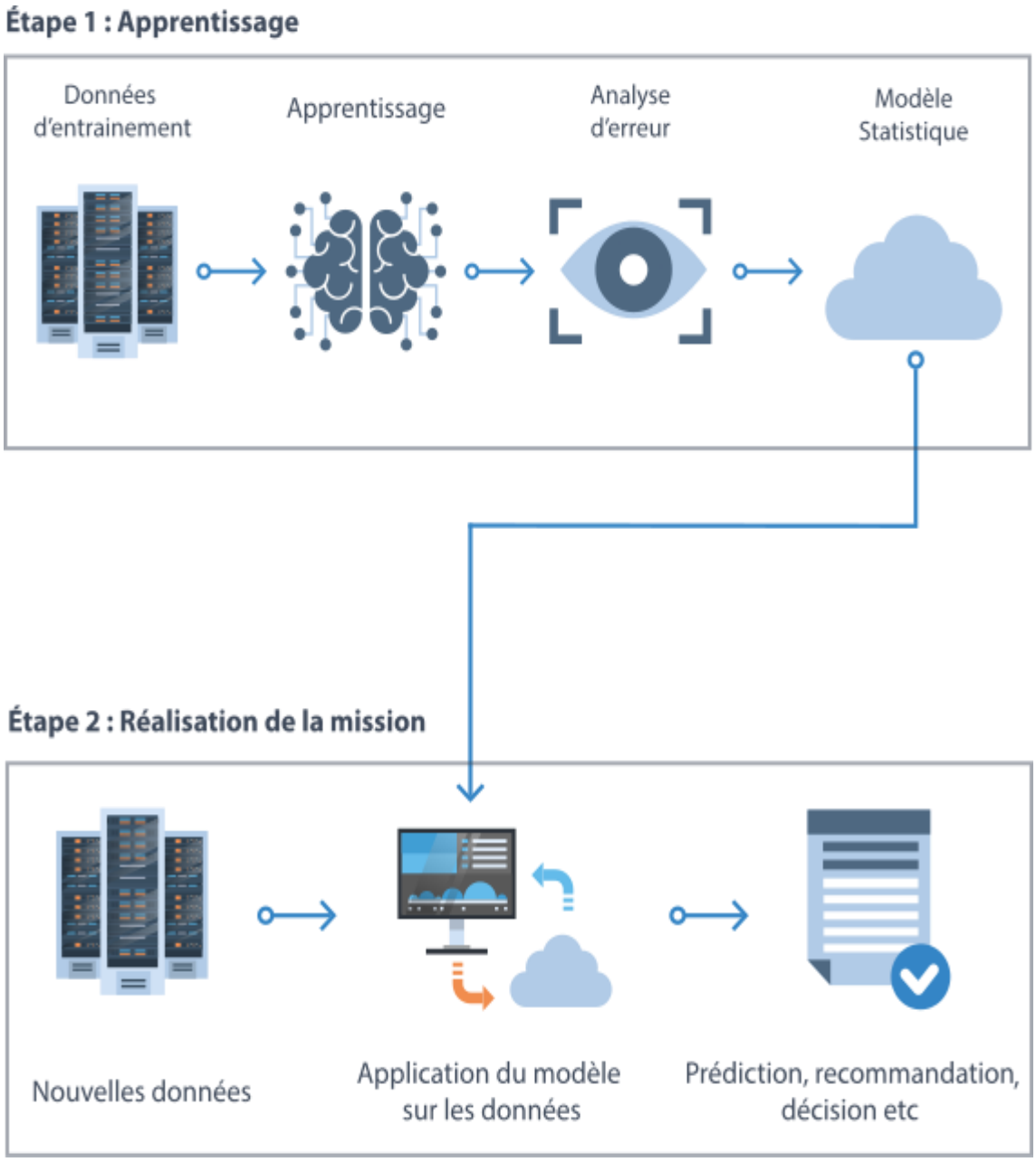

Figure 9 - Fonctionnement de la Machine Learning [18]

Par conséquent, le robot collectera beaucoup de données. C'est précisément à cause de cela qu'il peut devenir « intelligent ». Bien que le concept de « Machine Learning » ne soit pas réellement apparu récemment, il a été révélé en raison de l'arrivée du « Big Data ».

3.3.4 Les Avantages du machine learning ou apprentissage automatique

L'apprentissage automatique peut effectuer des tâches rapidement et parfois plus facilement que les humains, en particulier lorsqu'il s'agit d'une grande quantité de données. Ensuite, l'apprentissage automatique peut améliorer les performances, gagner du temps et augmenter l'efficacité.

3.3.5 Le Machine Learning et la Cybersécurité

Les cyberattaques se multiplient et les données à traiter se multiplient. Les experts en sécurité informatique risquent de ne pas être en mesure de détecter ces menaces avec succès. Ensuite, l'apprentissage automatique devient le principal atout pour détecter et gérer ces risques. En effet, l'un des plus grands défis auxquels sont confrontés les experts en cyber sécurité est d'anticiper les attaques de demain. S'il est aujourd'hui facile de détecter ou de prévenir les attaques connues, comment réagissez-vous aux nouveaux types d'attaques ? C'est là qu'entre en jeu le Machine Learning.

Dans notre plateforme les données massives collectées sur les interactions des utilisateurs et leurs statistiques dans les challenges et les simulations sont stockée et utilisée pour définir un modèle de Machine Learning pour créer des systèmes de recommandations et les utilisée encore à faciliter au chatbot d'interagir et aider les utilisateurs.

3.4 Le Deep Learning ou Apprentissage profond

La notion d'apprentissage profond est tout d'abord une traduction directe du terme anglais « Deep Learning », que certain préfère traduire par la notion d'apprentissage statistique. De même que sa traduction, sa définition varie également, mais principalement au niveau des détails. Pour définir cette notion dans les grandes lignes, on pourrait dire que :

L'apprentissage profond est un algorithme d'abstraction de haut niveau qui permet de modéliser les données à partir de grands ensembles de données apprises. [19]

3.4.1 Pourquoi profond ?

A l'origine, ces systèmes portaient le nom de réseaux artificiels de neurones (ANN, Artificial Neural Networks) afin de les différencier des systèmes biologiques. Ils se composent en général d'un certain nombre de données d'entrées et de sorties (input / output layer), d'un réseau étroit de neurones et de plusieurs strates intermédiaires (Hidden Layers). Ces couches intermédiaires permettent de traiter des problèmes complexes ; sans elles, le système ne résout que des calculs simples. Le nombre de couches est donc un facteur décisif pour la complexité du système, et de l'apprentissage ; les données s'associent d'une couche à l'autre, les résultats d'une première couche servant d'entrée à la prochaine, et ainsi de suite afin d'aboutir à une prise de décision complexe. Ce fonctionnement en strates donne toute sa profondeur au réseau et à l'apprentissage. L'adjectif « profond » s'entend ici dans tous les sens du terme. [19]

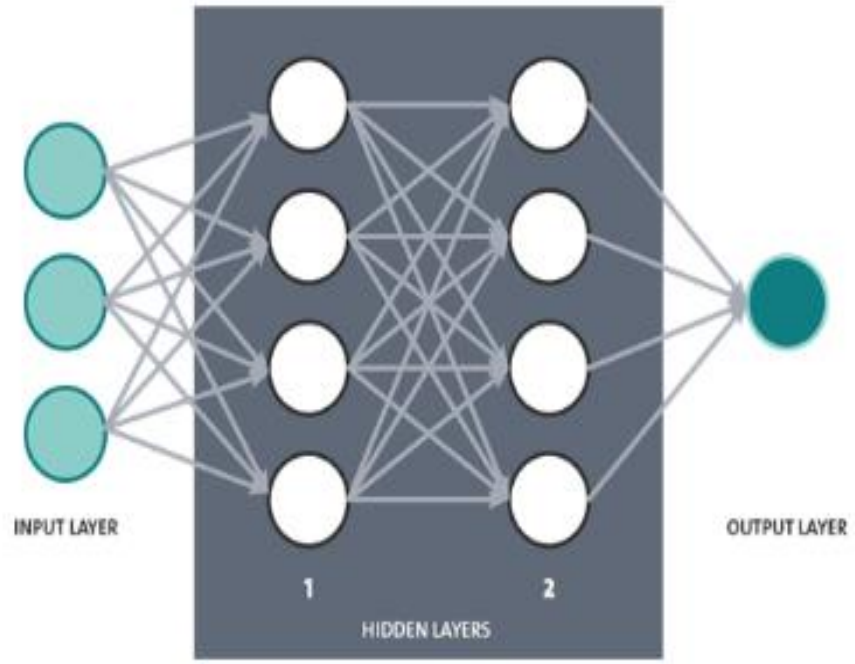

Figure 10 - Réseau de neurones [19]

L'exemple ci-dessus **(Voir Figure 10)** présente trois unités d'entrées, une sortie et deux couches intermédiaires (couches cachées). Vous constatez que les neurones sont « fortement interconnectés », ce qui constitue une propriété essentielle des réseaux de neurones. C'est justement ce qui autorise les relations, fonctions ou décisions d'être complexes, sans cette propriété, les relations entrée-sortie seraient relativement simples. [19]

Maintenant nous avons vu que l'apprentissage profond utilise les réseaux artificiels de neurones et on a vu brièvement c'est quoi un réseau artificiel de neurone, dans la section suivante nous allons voir c'est quoi exactement un réseau de neurone, ces types et quoi il est utilisée exactement, et comment ce dernier peut nous aider à créer nos systèmes intelligents dans notre plateforme.

3.5 Les réseaux artificiels de neurones

3.5.1 Qu'est-ce qu'un réseau de neurones ?

Tout d'abord, le réseau de neurones est un concept. Ce n'est pas physique. Le concept de réseaux de neurones artificiels (Artificial Neural Networks ANN) a été inspiré par les neurones biologiques. Dans un réseau de neurones biologiques, plusieurs neurones travaillent ensemble, reçoivent des signaux d'entrée, traitent des informations et déclenchent un signal de sortie.

Les réseaux de neurones tirent profit des expériences passées. [20]

3.5.3 Réseau de neurones d'Intelligence artificielle (IA)

Les réseaux de neurones artificiels et les réseaux de neurones biologiques sont basés sur le même modèle. **(Voir Figure 11)**

Bien que le concept sous-jacent soit le même que celui des réseaux biologiques, le réseau de neurones de l'IA est un groupe d'algorithmes mathématiques produisant une donnée de sortie (output) à partir des données d'entrée (input). [20]

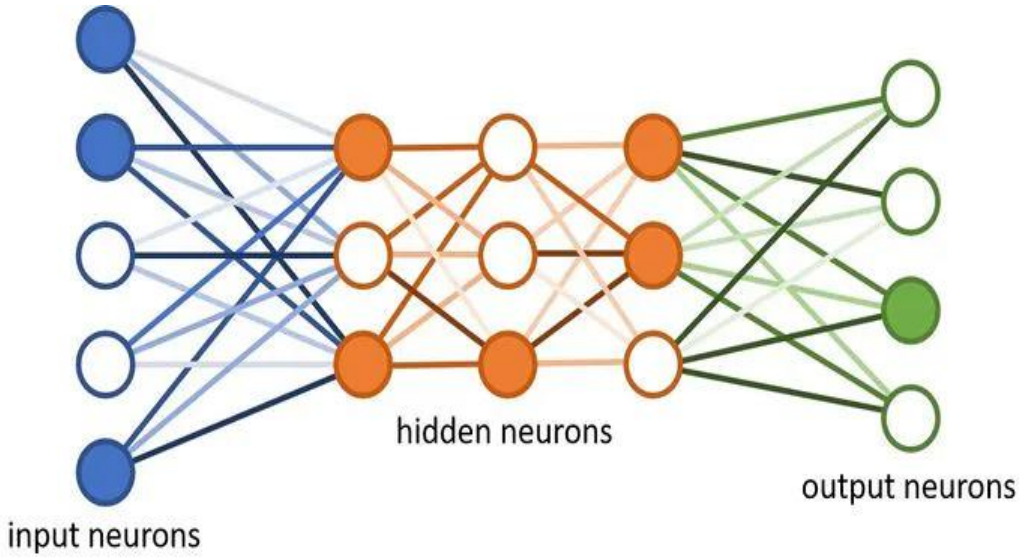

Figure 11 - Les réseaux de neurones artificiels [20]

Les réseaux de neurones sont entrainés avec une multitude de données d'entrées couplée à leurs données de sortie respectives. Ils calculent ensuite la donnée de sortie, ils la comparent à la donnée de sortie réelle connue et se mettent à jour en permanence pour améliorer les résultats (si nécessaire). [20]

En Résumé le réseau de neurones est un concept d'apprentissage automatique modélisé par le cerveau biologique.

3.5.3 Les composants d'un réseau de neurones

Le réseau de neurones est composé des composants principaux qui vont être expliquée dans les sous-sections qui suivent.

3.5.3.a Un neurone Artificiel, Qu'est-ce que c'est ?

L'idée de neurone artificiel n'a pas encore été vraiment spécifiée. Au sein d'un processeur, l'unité logique se compose de transistors ; on pourrait également y trouver un réseau neuronal « câblé », mais il faudrait qu'il soit « adaptatif », qu'il ait une « capacité d'apprentissage ». En effet, la réponse d'un neurone à des impulsions entrantes doit pouvoir évoluer tout au long du processus d'apprentissage. C'est ce qu'on appelle la « pondération » : un neurone évalue (pondère) diverses variables d'entrée pour obtenir la variable de sortie souhaitée. C'est pourquoi les neurones sont généralement des fonctions mathématiques qui relient entre elles des variables d'entrée et de sortie. [19]

3.5.3.b Couches: groupement de neurones

Les couches (ou couches) contiennent des neurones et aident à diffuser les informations. Il y a au moins deux couches dans un réseau neuronal: la couche d'entrée et la couche de sortie. Il est très probable qu'il existe un (très) grand nombre de couches dans un réseau neuronal complexe. Plus il y a de couches, plus le réseau est profond (apprentissage en profondeur). Les couches autres que les couches d'entrée et de sortie sont appelées couches cachées.

3.5.3.c Poids et biais: valeurs numériques

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

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

Les biais sont également des valeurs numériques qui sont ajoutées une fois que les poids sont appliqués aux valeurs d'entrée. Les poids et les biais sont donc en quelque sorte des valeurs d'auto-apprentissage de nos réseaux de neurones.

Considérez le poids comme une donnée capitale pour un neurone.

3.5.3.d Fonction d'activation

La fonction d'activation est tout simplement un algorithme mathématique appliqué aux valeurs de sortie d'un réseau de neurone.

Il existe un grand nombre de fonctions d'activation, telles que [20] :

Sigmoïde: produit une courbe en forme de S. Bien que de nature non linéaire, il ne tient toutefois pas compte des légères variations des entrées, ce qui entraîne des résultats similaires.

Fonctions de tangente hyperbolique (**tanh**): Il s'agit d'une fonction supérieure comparée à Sigmoïde. Cependant, elle rend moins bien compte des relations et elle est plus lente à converger.

Unité linéaire rectifiée (**Relu**): Cette fonction converge plus rapidement, optimise et produit la valeur souhaitée plus rapidement. C'est de loin la fonction d'activation la plus populaire utilisée dans les couches cachées.

Softmax: utilisé dans la couche de sortie car il réduit les dimensions et peut représenter une distribution catégorique.

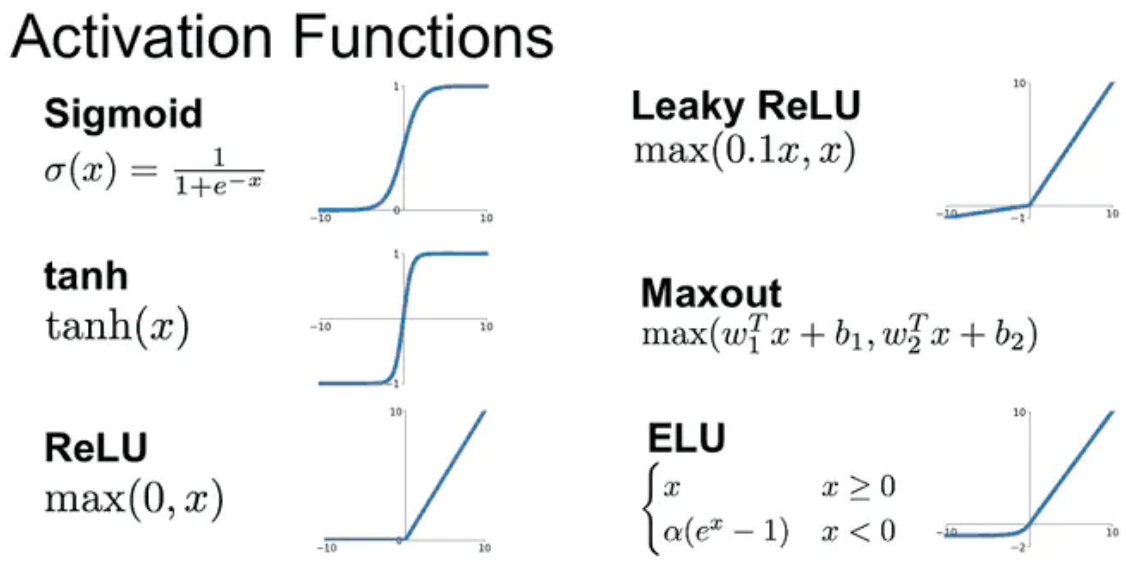

Figure 12 - Les fonctions d'activation [20]

3.5.4 Fonctionnement des réseaux de neurones

Le concept de réseaux de neurones repose sur trois étapes principales:

- Pour chaque neurone de la couche, la valeur d'entrée est multipliée par le poids.
- Ensuite, pour chaque couche, ajoutez tous les poids des neurones et ajoutez le biais.
- Enfin, la fonction d'activation est appliquée à la valeur pour calculer la nouvelle sortie.

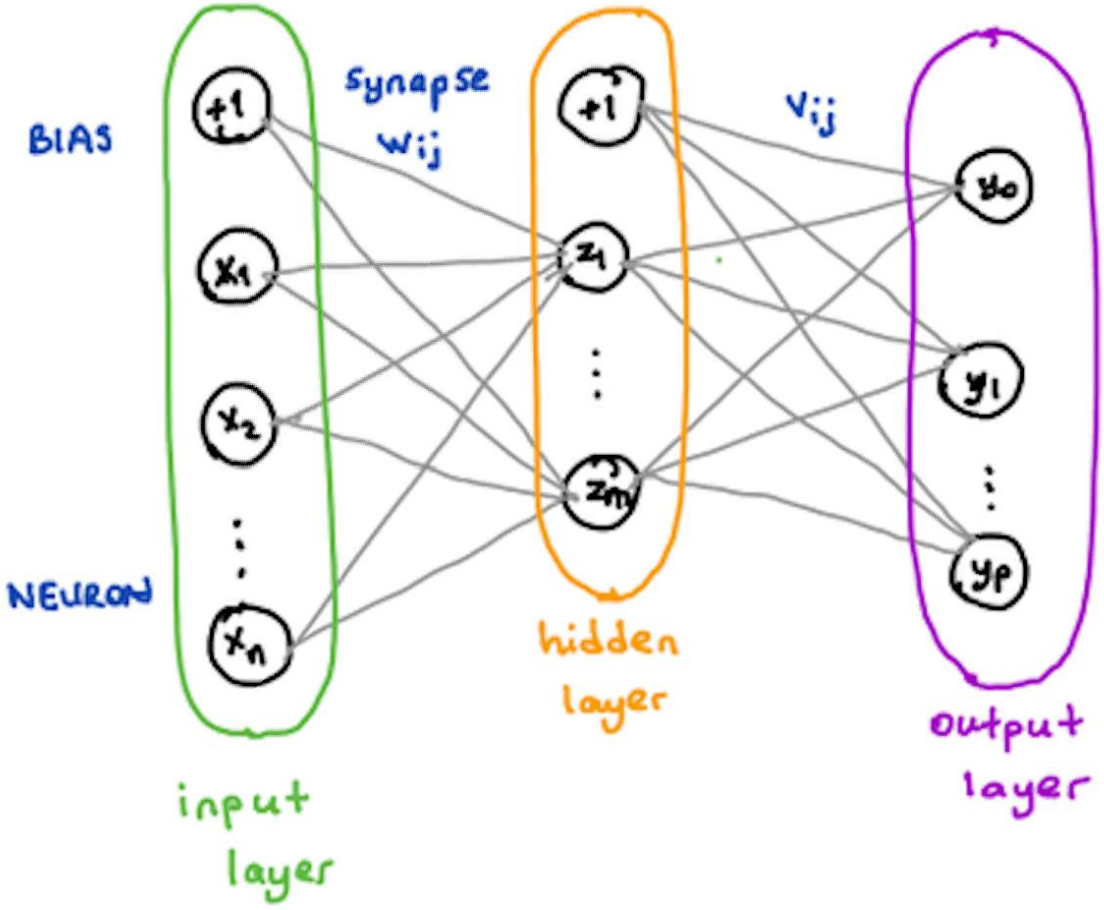

Figure 13 - Fonctionnement des réseaux de neurones [19]

3.5.5 Les types des réseaux de neurones

Il existe différents types de réseaux de neurones. Les deux réseaux de neurones les plus populaires sont citées dans les sous-sections suivantes.

3.5.5.A Réseau de neurones récurrent – Recurrent Neural Network (RNN):

Ce sont des réseaux de neurones spécialisés qui utilisent le contexte des entrées lors du calcul de la sortie. La sortie dépend des entrées et des sorties calculées précédemment.

Ces réseaux nous aident à prévoir les séries chronologiques dans les applications commerciales et à prévoir les mots dans les applications de type chatbot. [20]

Ce type de réseau de neurone est important dans notre concept de chatbot, qui est comme citée précédemment utile pour aider les utilisateurs à se retrouver dans la plateforme et savoir comment l'utiliser.

La figure 14 illustre le fonctionnement d'un réseau de neurone récurrent

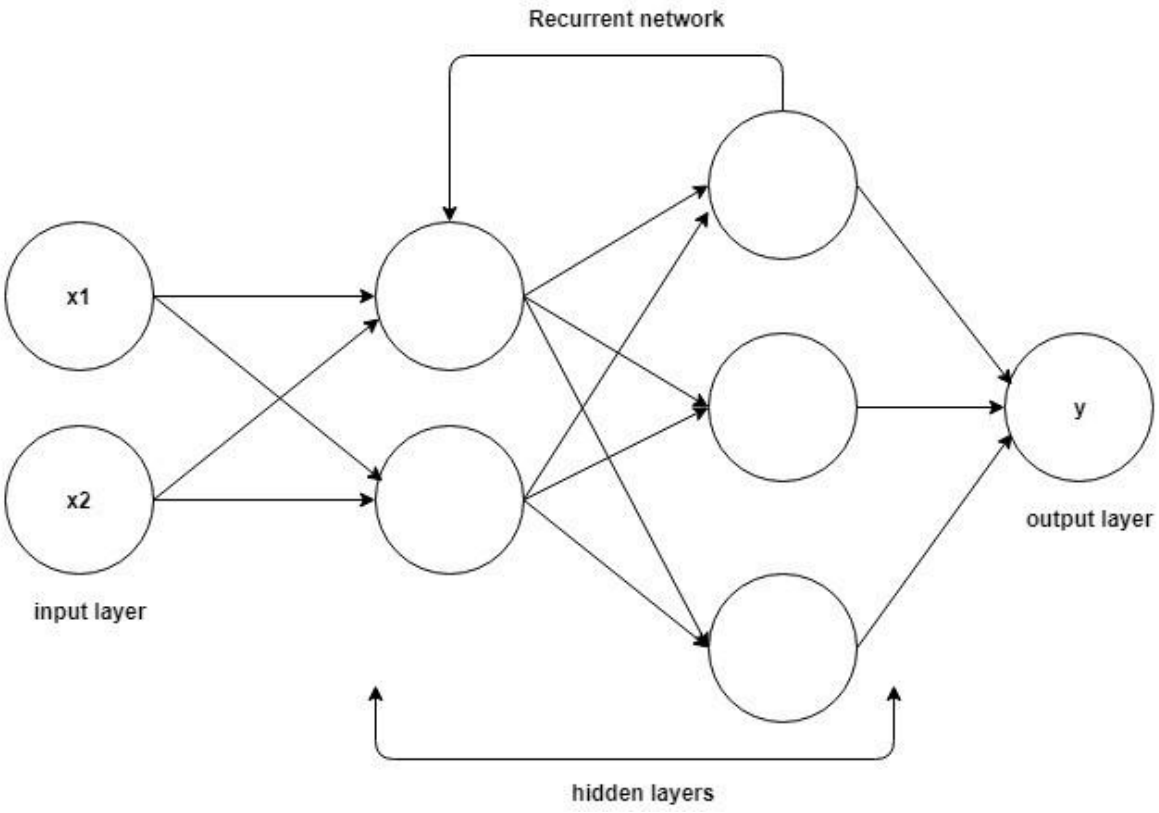

Figure 14 - Réseau de neurones récurrent [20]

3.5.5.b Réseau de neurones de convolution – Convolution Neural Network (CNN):

Ces réseaux reposent sur des filtres de convolution (matrices numériques). Les filtres sont appliqués aux entrées avant que celles-ci ne soient transmises aux neurones.

Ces réseaux de neurones sont utiles pour le traitement et la prévision d'images.

Pour notre plateforme de type est pas trop utilisée mais il est très important pour comprendre le raisonnement des systèmes de Deep Learning et de réseaux de neurone.

Une illustration du fonctionnement d'un Réseau de neurones de convolution est montrée sur la Figure 15.

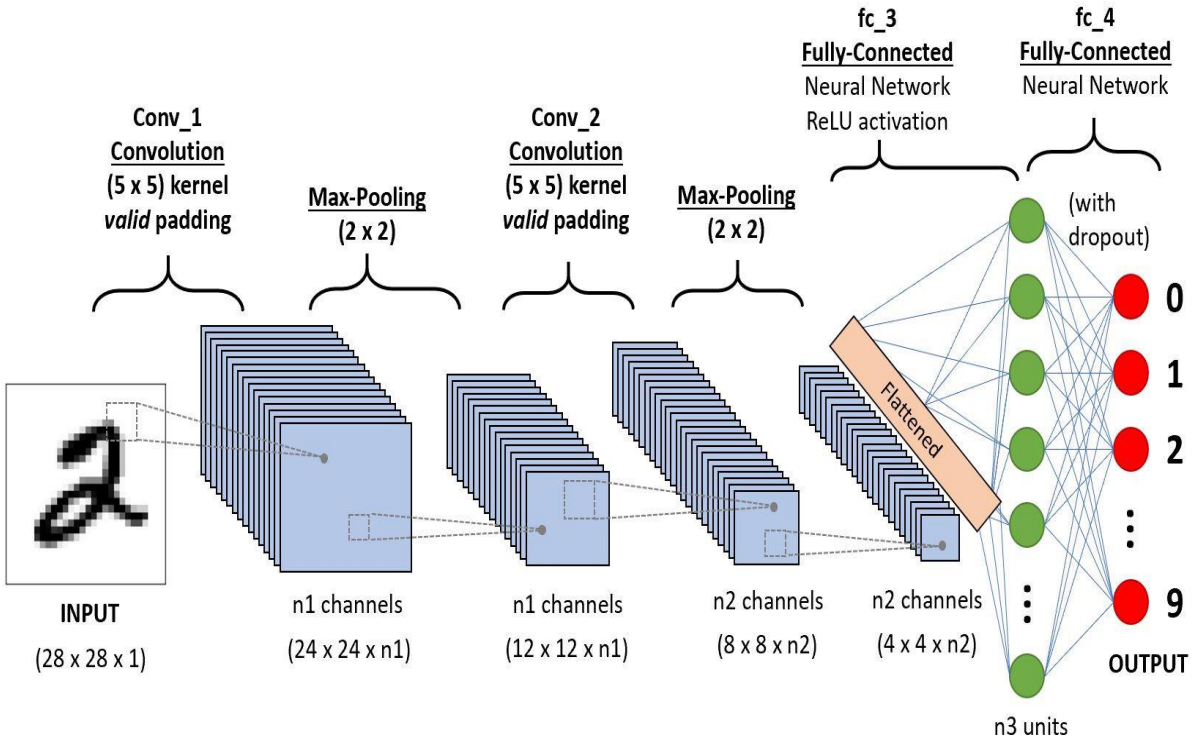

Figure 15 - Réseau de neurones de convolution [20]

Dans cette section on a expliqué brièvement c'est quoi les réseaux de neurone artificiel, leur fonctionnement, les principaux composants d'un réseau de neurone et enfin les types de ces derniers d'où on a vu que les réseaux de neurones récurrent sont très important à utiliser dans notre chatbot.

Donc, nous avons vu quelque généralité sur l'intelligence artificielle, l'apprentissage automatique et profond, les réseaux neurone qui sont important pour comprendre ce qui suit, Maintenant on va entamer l'essentiel qui est le principe de nos systèmes intelligent pour la plateforme, oui on parle des systèmes de recommandations intelligent et les concepts utilisée dans notre chatbot.

3.6 Chatbots: Théorie

3.6.1 Qu'est-ce que la conversation ?

La conversation est l'échange d'informations entre au moins deux personnes, généralement sur un sujet spécifique. La conversation est une forme courante de communication où les gens peuvent se comprendre et discuter sur un sujet spécifique, une conversation peut être mise en place pour poser des questions ou savoir des fonctionnements de certains concepts.

Comme dans notre système la compréhension de la plateforme ou des concepts peut nécessiter la mise en place d'une conversation, mais donner à un agent une responsabilité de répondre à des milliers de conversation peut être une tâche impossible, c'est pour ça le concept d'un chatbot est requis.

3.6.2 Qu'est-ce Qu'un chatbot ?

Un chatbot est un programme qui essaie de parler à quelqu'un pendant quelques minutes ou plus tout en lui donnant l'impression de parler avec une vraie personne. Une bonne compréhension de la conversation aidera à avoir une conversation significative, mais la plupart des chatbots n'essaieront pas de parler. En particulier, ils trouvent des mots dits déclencheurs et même les expressions de l'interlocuteur pour trouver des réponses. Les programmes de ces réponses sont programmés pour guider d'avantage la conversation de manière plus ou moins intelligente sans savoir de quoi ils parlent.

Par exemple si un utilisateur pose la question au chatbot « c'est quoi une injection SQL » le chatbot peut capter le mot « injection SQL » comme déclencheurs, ou un autre déclencheur qui est similaire a ce mot et invoque les réponses similaire et correspondant à ce déclencher.

3.6.3 Les Avantages d'un chatbot

Bien sûr on ne peut pas parler d'un concept sans parler de ses avantages, le chatbot contient beaucoup d'avantages, mais on va citer les plus importants.

Le chatbot en fin de compte c'est un programme numérique, et quand on dit numérique alors il y a des données stockées dans les ordinateurs et sur des bases de données, donc chaque conversation peut être enregistrée et stockée pour entrainer le chatbot ou encore

pour les utiliser dans des suivis ou dans des trucs marketing. Aussi un chatbot peut gérer une importante quantité des conversations et sur des sujets différents avec une rapidité extraordinaire et sans perdre trop de temps.

En conclusion, le bot est mis à disposition pour faire gagner du temps aux clients. La création d'un bot qui fixe des rendez-vous est bénéfique pour le client et les employés qui n'auront plus besoin d'attendre la réponse de l'entreprise ou de perdre du temps à chercher une plage horaire de libre qui conviendrait au client.

3.6.4 Chatbot sans intelligence artificielle

Les robots qui n'utilisent pas l'IA répondent aux utilisateurs à l'aide de commandes qui commencent par un «!» Suivi de l'information qui amène la réponse à la question voulue. Un des désavantages dans le fait de ne pas avoir d'intelligence artificielle, c'est qu'il est nécessaire de connaitre les commandes par cœur ou de chercher dans l'index proposer avec toutes les actions possibles. Dans les sites de streaming de jeux-vidéos, par exemple, des milliers de personnes écrivent en même temps dans le chat privé de leur streamer préféré. Il serait impossible que le bot arrive à suivre la conversation et réponde à tout le monde dans un temps record. Mais encore, la majorité des messages concerne principalement la partie du jeu que le streamer est en train de jouer. C'est pour cela que les commandes sont utiles pour répondre aux questions qui sont le plus souvent posées par les fans, on pourrait comparer cela à une page FAQ d'une entreprise. [24]

3.6.5 Chatbot Avec Intelligence Artificielle

L'utilité de parler à un agent conversationnel, c'est qu'il va essayer de comprendre l'intention et d'y répondre. C'est une discussion entre humain et machine, le créateur aimerait rendre la machine aussi vivante qu'un être humain. [24]

Les humains veulent parler aux machines comme ils se parlent entre eux, c'est à dire en langage « naturel », pas dans un langage de machine. Le champ du NLP (Natural Language Processing ou TLN en français) a justement émergé pour répondre aux besoins des machines d'interpréter des informations formulées en langage naturel.

En d'autres termes, le NLP permet aux humains et aux machines de se parler avec une syntaxe et un vocabulaire qui est celui des hommes. Des systèmes NLP ont pour vocation de comprendre ce que les humains disent, de traiter la donnée qui est dans le message et, si besoin, d'agir avant de donner une réponse - elle aussi en langage naturel. [25]

3.6.5.a NLP (Natural Language Processing)

Le NLP (Natural Language Processing) est une branche de l'intelligence artificielle spécialisée dans le traitement du langage écrit, également connu sous le nom français TALN (traitement automatique du langage naturel).

La NLP permet aux humains et aux machines de communiquer entre eux en utilisant la grammaire et le vocabulaire humains. Les systèmes NLP sont conçus pour comprendre ce que les gens disent, traiter les données du message et prendre des mesures avant de répondre si nécessaire - également en langage naturel.

Transformation de base

Comme mentionné précédemment, pour qu'une machine donne un sens au langage naturel (langage utilisé par les humains), elle doit être convertie en une sorte de cadre mathématique qui peut être modélisé. Vous trouverez ci-dessous quelques-unes des techniques les plus couramment utilisées qui nous aident à atteindre cet objectif. [27]

Méthodes de transformation

L'une des méthodes les plus courantes pour y parvenir dans un sac de représentation de mots est tf-idf

TF-IDF

TF-IDF est une manière de noter le vocabulaire afin de donner un poids adéquat à un mot en proportion de l'impact qu'il a sur le sens d'une phrase. Le score est un produit de 2 scores indépendants, la fréquence des termes (TF) et la fréquence inverse des documents (IDF). [27]

La fonction mathématique de TF-IDF est illustrée sur la figure 16.

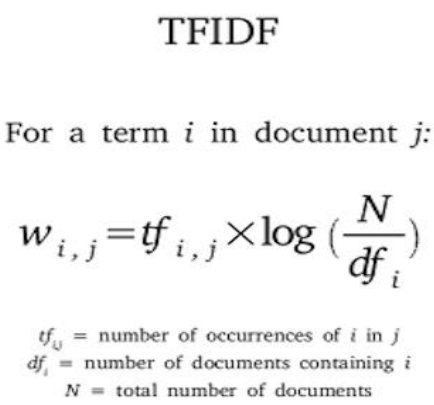

Figure 16 - Fonction TFIDF [27]

Encodages One-Hot

Les encodages à chaud sont une autre façon de représenter les mots sous forme numérique. La longueur du vecteur de mot est égale à la longueur du vocabulaire, et chaque observation est représentée par une matrice avec des lignes égales à la longueur du vocabulaire et des colonnes égales à la longueur de l'observation, avec une valeur de 1 où le mot du vocabulaire est présent dans l'observation et une valeur de zéro là où ce n'est pas le cas.

La figure 17 est une illustration de l'Encodages One-Hot.

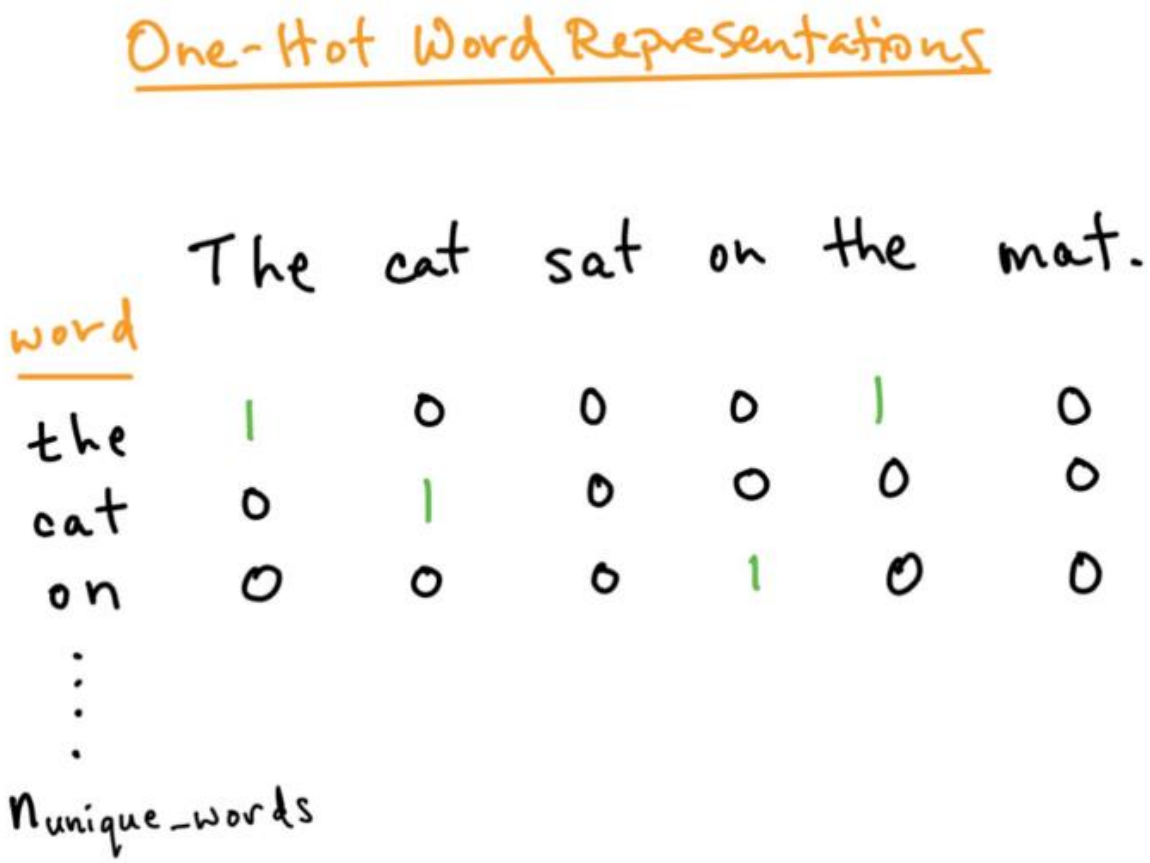

Figure 17 - illustration de l'encodage One-Hot [27]

Word Embeddings

Word Embeddings est le nom collectif d'un ensemble de techniques de modélisation du langage et d'apprentissage de fonctionnalités où des mots ou des phrases du vocabulaire sont mappés à des vecteurs de nombres réels. La technique est principalement utilisée avec les modèles de réseau neuronal.

La figure 18 illustre le fonctionnement de la technique de « Word Embeddings ».

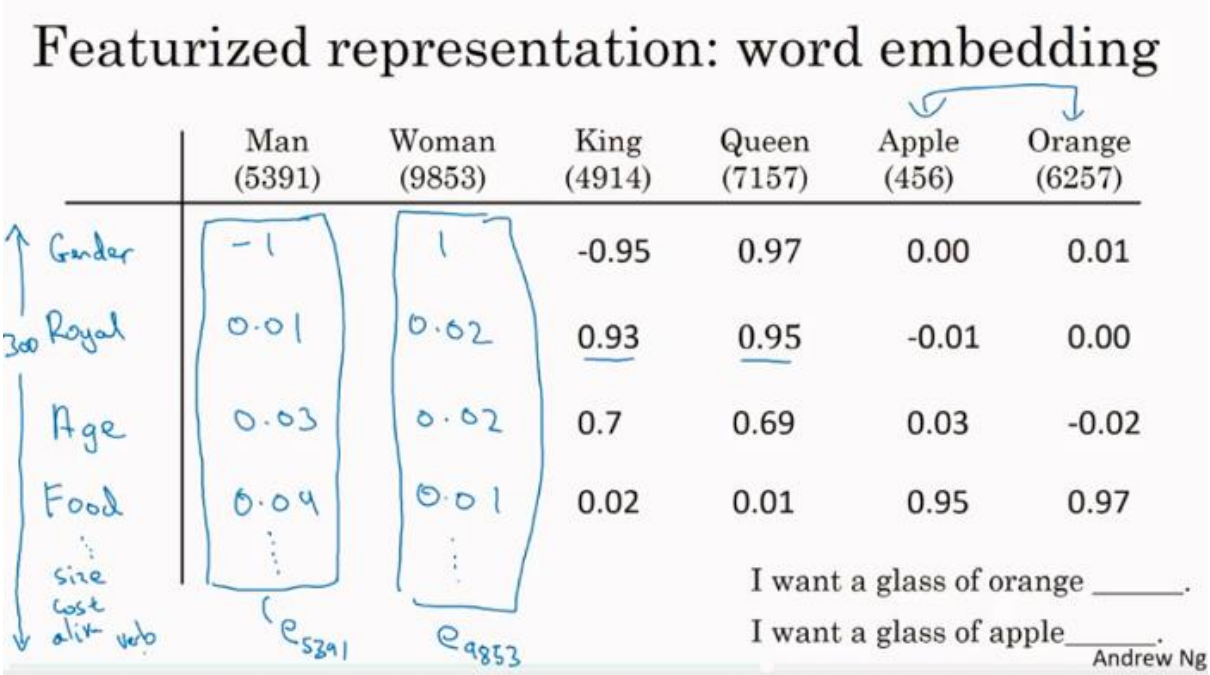

Figure 18 - le fonctionnement de la technique de « Word Embeddings » [27]

3.6.5.b NLU (Natural Language Understanding)

La compréhension du langage naturel (Natural Language Understanding ou NLU en anglais) permet aux chatbots de comprendre le langage humain non structuré.

3.6.5.C NLP et NLU

Le NLU est la sous-partie du NLP qui se concentre principalement sur l'amélioration de la capacité des machines à comprendre la signification qui se cache derrière les mots et dans les textes.

Pour résumer, la différence entre le NLP et le NLU tient au fait que le premier s'attache à interpréter littéralement ce que les humains disent ou écrivent, là où le deuxième s'attache à identifier les intentions et la signification profonde de ce qui est dit ou écrit. [25]

La figure 19 illustre la différence NLP, NLU et leur rapport avec l'intelligence artificielle.

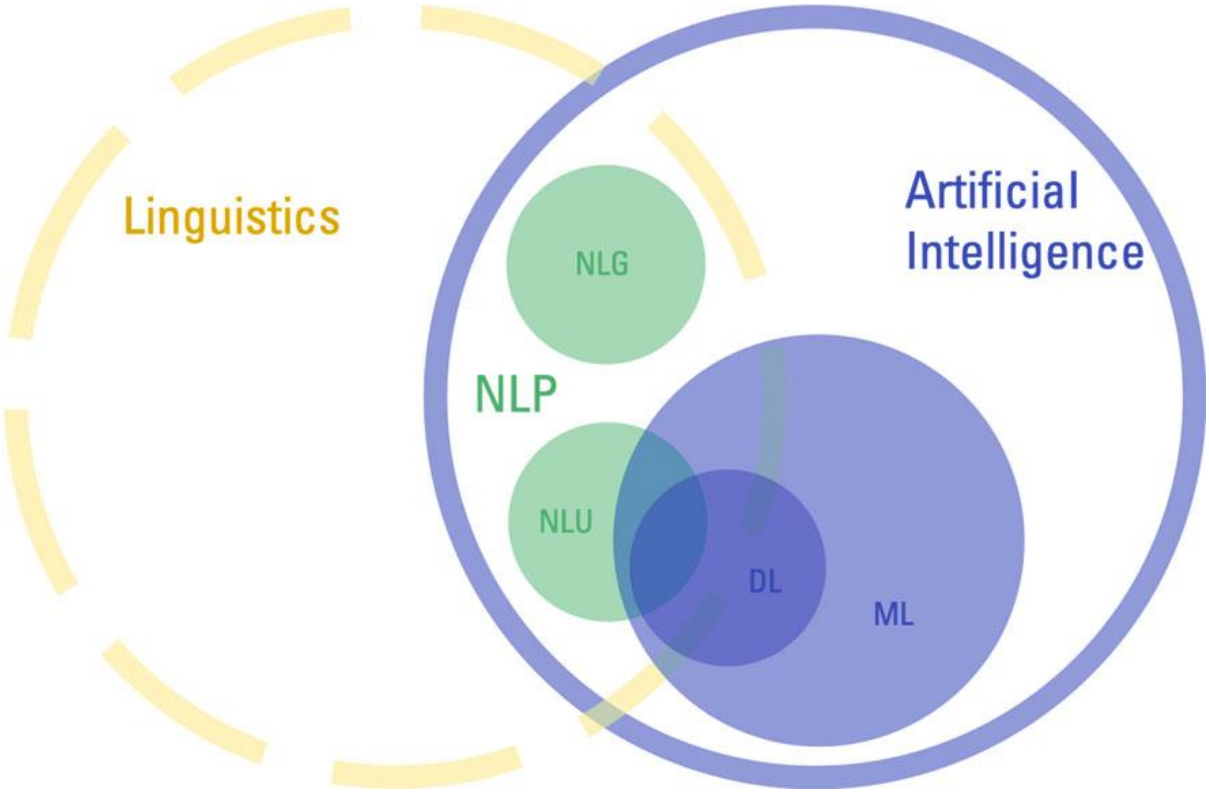

Figure 19 - la différence NLP, NLU [25]

Les réseaux de neurones récurrents ou RNN comme on les vu dans une section précédente, sont une variante très importante des réseaux de neurones largement utilisés dans le traitement du langage naturel (NLP).

3.7 Les systèmes de recommandations Intelligents

On peut appliquer des systèmes de recommandation dans des scénarios où de nombreux utilisateurs interagissent avec de nombreux éléments, dans notre exemple les scénarios c'est les interactions des utilisateurs et leur avis dans les différents challenges et machines, et encore leur accès pour donner des recommandations aux utilisateurs sur les formations et les challenges qui peuvent être aimée par l'utilisateur en question.

3.7.1 Pourquoi les recommandations ?

Un système de recommandation aide les utilisateurs à trouver un contenu convaincant dans un grand nombre d'objets. Par exemple, le Google Play Store fournit des millions d'applications, tandis que YouTube fournit des milliards de vidéos. Plus d'applications et de vidéos sont ajoutées chaque jour. Comment les utilisateurs peuvent-ils trouver de nouveaux contenus intéressants? Oui, on peut utiliser la recherche pour accéder au contenu. Cependant,

un moteur de recommandation peut afficher des éléments que les utilisateurs n'auraient peut-être pas pensé rechercher eux-mêmes.

3.7.2 Terminologie

Éléments (également appelés documents)

C'est les entités qu'un système recommande. Pour le Google Play Store, les éléments sont des applications à installer. Pour YouTube, les éléments sont des vidéos.

Les éléments dans notre cas c'est les formations à s'inscrire ou les challenges a essayer.

Requête (également appelée contexte)

Les informations qu'un système utilise pour faire des recommandations.

Dans notre plateforme c'est l'identifiant des utilisateurs avec ces avis sur les challenges ou encore les accès de l'utilisateur.

Embedding

C'est la transformation ou le mappage d'un ensemble discret (dans ce cas, l'ensemble des requêtes, ou l'ensemble des éléments à recommander) à un espace vectoriel appelé « Embedding Space ».

3.7.3 Les composants d'un système de recommandation

Génération de candidats (Candidate Generation)

Dans la première étape, le système démarre avec un nombre de données potentiellement énorme et génère un sous-ensemble plus restreint de candidats. Par exemple, le générateur de candidats de YouTube réduit des milliards de vidéos à des centaines ou des milliers. Compte tenu d'un corpus important, le modèle doit évaluer rapidement les demandes. Un modèle donné peut fournir plusieurs générateurs de candidats, dont chacun spécifie un sous-ensemble différent de candidats.

Par exemple dans notre cas le générateur de candidats de notre plateforme peut réduire des milliers des challenges et des formations a des certaines.

Ce composant est très important dans les systèmes de recommandations

Notation (Scoring)

Ensuite, un autre modèle classe et classe les candidats afin de sélectionner l'ensemble des éléments à afficher à l'utilisateur.

Reclassement (Re-ranking)

Enfin, le système doit tenir compte d'autres limitations de la classification finale. Par exemple, le système supprime les éléments que l'utilisateur n'aime pas clairement ou augmente le score du nouveau contenu.

3.7.4 Classification des systèmes de recommandations

Les systèmes de recommandation peuvent être classés en 3 types qui sont utilisée aujourd'hui dans les systèmes informatiques.

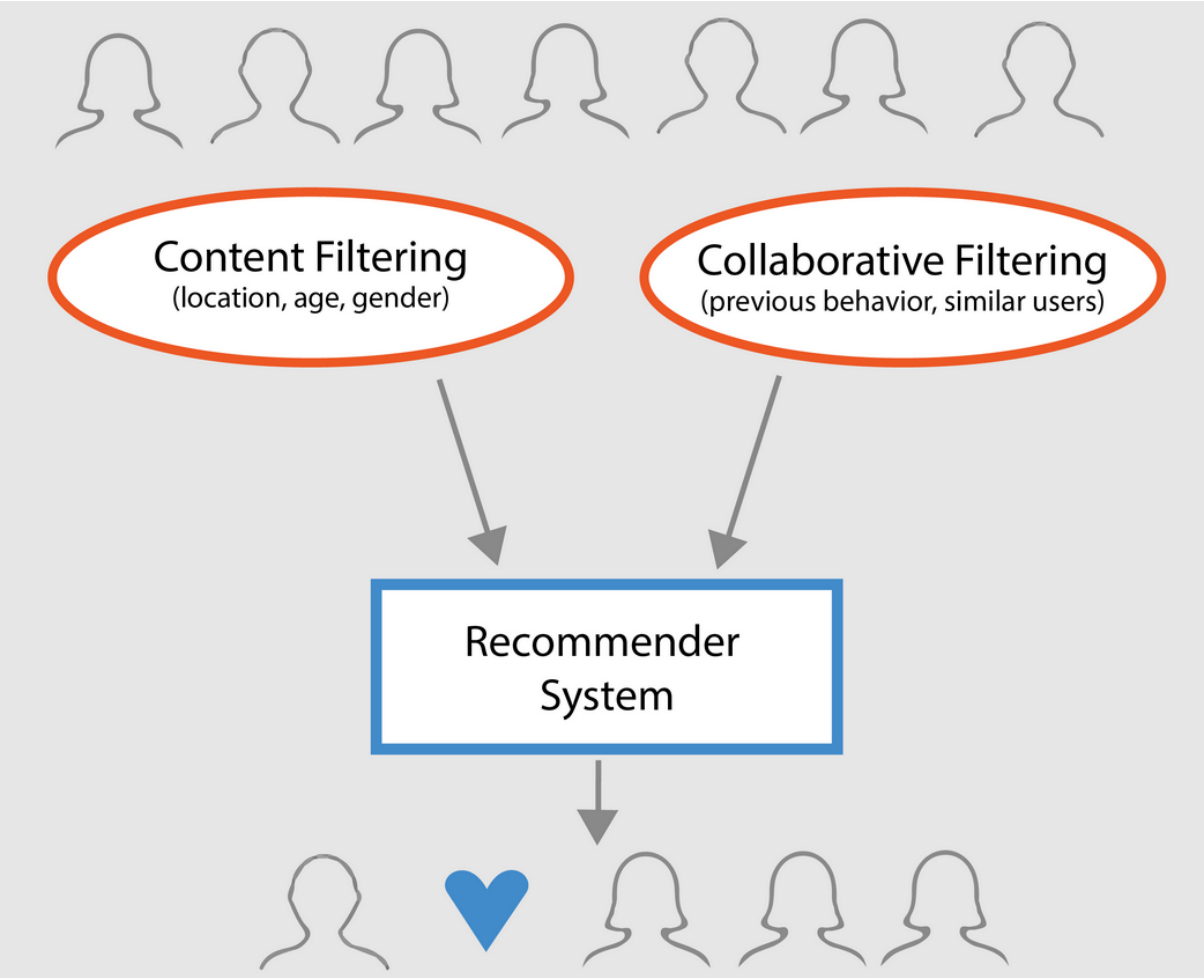

Figure 20 - Classification des systèmes de recommandation

3.7.4.A Recommandation simple

Propose des recommandations généralisées à chaque utilisateur, basées sur les avis ou les catégories des formations et des challenges. L'idée de base derrière ce système est que les catégories les plus populaires et les plus aimées par les utilisateurs auront une probabilité plus élevée d'être appréciés par le public moyen.

Exemple : La catégorie « Web Challenge » a des avis de 5 étoiles par les utilisateurs, et la catégorie « Cryptographie » a des avis de 4 étoiles, donc la catégorie « Web Challenge » a une probabilité plus élevée d'être appréciés.

3.7.4.b Les recommandations basées sur le contenu (Content based)

Suggère des challenges similaires en fonction d'un challenge particulier. Ce système utilise des métadonnées d'éléments, telles que la catégorie, l'auteur, la description, etc. pour les formations ou les challenges, pour faire ces recommandations. L'idée générale derrière ces systèmes de recommandation est que si une personne aime un challenge particulier, elle aimera également un challenge qui lui est similaire et une formation de même catégorie. Et pour le recommander, il utilisera les métadonnées des éléments passés de l'utilisateur. Un bon exemple pourrait être YouTube, où, en fonction de votre historique, il vous suggère de nouvelles vidéos que vous pourriez potentiellement regarder. **(Voir Figure 21)**

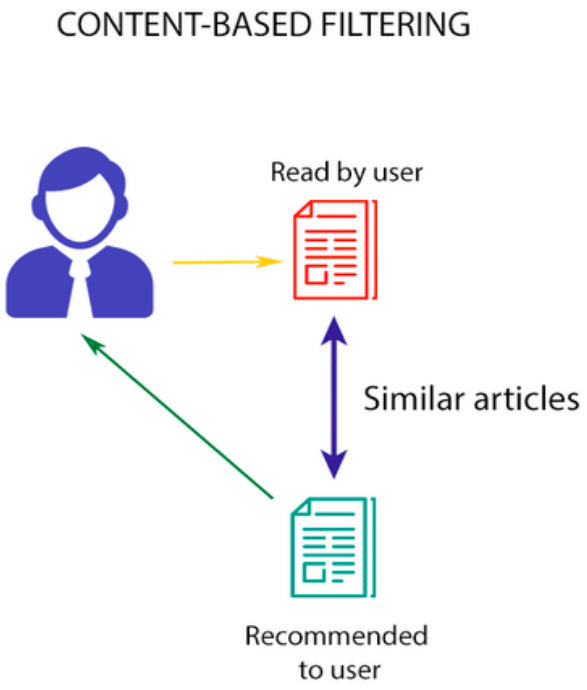

Figure 21 - Les recommandations basées sur le contenu. [28]

Exemple : Un utilisateur à plusieurs accès sur les catégories des challenges web et particulièrement aux challenges de l'injection SQL, donc il peut aimer des challenges de même catégorie, ou que leur description ou titres contient « SQL Injection ».

Les Mesures et techniques utilisée

Mesures de similarité

Une mesure de similarité est une fonction $E \times E \rightarrow \mathbb{R}$ qui prend une paire d'incorporations et renvoie un scalaire mesurant leur similarité.

Pour déterminer le degré de similitude, la plupart des systèmes de recommandation reposent sur un ou plusieurs des éléments suivants

Cosine

C'est simplement le cosinus de l'angle entre les deux vecteurs $s(q,x) = \cos(q,x)$

Produit scalaire

Le produit scalaire de deux vecteurs est $s(q,x) = (q,x) = \sum_{i=1}^{d} q_i\, x_i$. Il est également donné par $s(q,x) = \|x\|\|q\| \cos(q,x)$ (le cosinus de l'angle multiplié par le produit des normes). Ainsi, si les plongements sont normalisés, le produit scalaire et le cosinus coïncident.

Ces mesures et techniques sont utilisées dans notre plateforme pour les recommandations basées sur le contenu.

Avantage

- Le modèle n'a pas besoin de données sur les autres utilisateurs, car les recommandations sont spécifiques à cet utilisateur. Cela facilite l'adaptation à un grand nombre d'utilisateurs.

- Le modèle peut capturer les intérêts spécifiques d'un utilisateur et peut recommander des éléments de niche qui intéressent très peu d'autres utilisateurs.

Inconvénients

- Étant donné que la représentation des caractéristiques des éléments est conçue à la main dans une certaine mesure, cette technique nécessite beaucoup de connaissances du domaine. Par conséquent, le modèle ne peut être aussi bon que les fonctionnalités conçues à la main.

- Le modèle ne peut faire que des recommandations basées sur les intérêts existants de l'utilisateur. En d'autres termes, le modèle a une capacité limitée à développer les intérêts existants des utilisateurs.

3.7.4.c Les recommandations de filtrage collaboratif (Collaborative filtering)

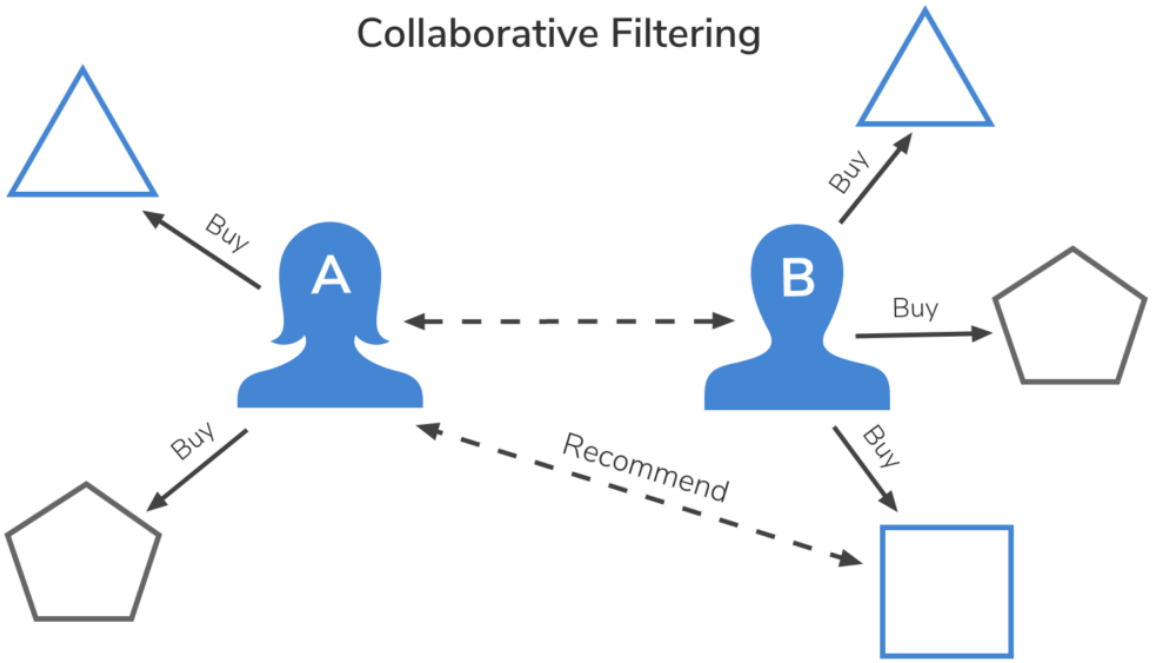

Figure 22 - Les recommandation de filtrage collaboratif. [28]

Ces systèmes sont largement utilisés et tentent de prédire la note ou la préférence d'un utilisateur pour un challenge en fonction des notes passées et des préférences d'autres utilisateurs. Les filtres collaboratifs ne nécessitent pas de métadonnées de projet comme le contenu basé sur le contenu.

Par exemple si un utilisateur **A** aimé « SQL injection », « XSS » et l'utilisateur **B** aime aussi « SQL injection », XSS et il aime « LFI » donc il y a une grande probabilité que l'utilisateur A peut aimer « LFI ».

Modèle utilisée

Factorisation matricielle [27]

La factorisation matricielle est un modèle d'intégration simple. Compte tenu de la matrice des avis $A \in \mathbb{R}^{m \times n}$, où m est le nombre d'utilisateurs (ou de requêtes) et $n\ est$ le nombre d'éléments, le modèle apprend :

- Une matrice d'intégration utilisateur $U \in \mathbb{R}^{m \times d}$, où la ligne i est l'intégration pour l'utilisateur i.
- Une matrice d'intégration d'élément $V \in \mathbb{R}^{n \times d}$, où la ligne j est l'incorporation de l'élément j.

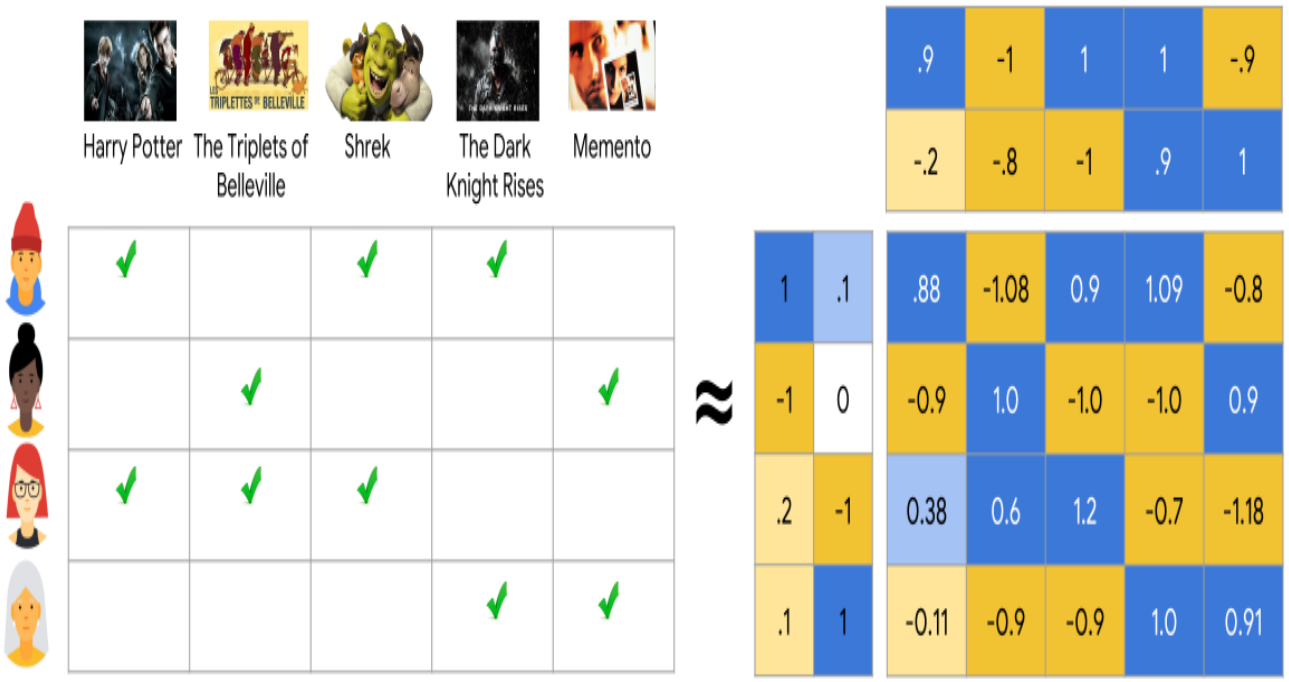

Figure 23 - Factorisation Matricielle. [27]

Le modèle apprend de telle sorte que le produit UV^T soit une bonne approximation de la matrice des avis A. Observez que (i, j) l'entrée de $U.V^T$ est simplement le produit scalaire (U_i, V_j) des plongements de l'utilisateur i et de l'élément j, dont vous voulez que il soit proche de $A_{i,j}$.

Avantage

- Nous n'avons pas besoin de connaissances du domaine car les incorporations sont automatiquement apprises.

- Le modèle peut aider les utilisateurs à découvrir de nouveaux intérêts. De manière isolée, le système ML peut ne pas savoir que l'utilisateur est intéressé par un élément donné, mais le modèle peut toujours le recommander car des utilisateurs similaires sont intéressés par cet élément.

- Dans une certaine mesure, le système n'a besoin que de la matrice de rétroaction pour former un modèle de factorisation matricielle. En particulier, le système n'a pas besoin de fonctionnalités contextuelles. En pratique, cela peut être utilisé comme l'un des multiples générateurs candidats.

Inconvénients

- La prédiction du modèle pour une paire (utilisateur, élément) donnée est le produit scalaire des plongements correspondants. Ainsi, si un élément n'est pas vu pendant la formation, le système ne peut pas créer une incorporation pour lui et ne peut pas interroger le modèle avec cet élément. Ce problème est souvent appelé problème de démarrage à froid.

- Les fonctionnalités secondaires sont toutes les fonctionnalités au-delà de la requête ou de l'ID d'élément. Pour les recommandations de films, les fonctionnalités secondaires peuvent inclure le pays ou l'âge. L'inclusion des fonctionnalités latérales disponibles améliore la qualité du modèle. Bien qu'il ne soit pas facile d'inclure des fonctionnalités secondaires dans WALS, une généralisation de WALS rend cela possible.

3.7.5 Dataset et choix de la classe de recommandation

3.7.5.a Jeu de données (Dataset)

Le jeu de données est un grand nombre de données généré à travers les programmes ou peut être trouvé sur internet selon le thème, mais dans le domaine de la sécurité informatique les datasets sont rare et spécialement dans les systèmes de recommandations en rapport avec la sécurité informatique ou même d'autres trucs, **pour ça l'un des buts de notre plateforme est de générer et créer ce genre de dataset qui peuvent être utilisée dans les recherches, les systèmes de recommandations ouvert etc. …**

Dans la plateforme le premier jeu de données c'est les avis des utilisateurs sur les différentes catégories de la sécurité informatique et leurs types qui sont générés à partir des avis des clients sur les challenges.

Un deuxième jeu de données important est l'accès des utilisateurs aux challenges, veut dire de suivre les accès des utilisateurs sur les différentes catégories et types dans les challenges, par exemple l'utilisateur **A** la majorité de ces accès sont sur les challenges de type « SQL Injection ».

3.7.5.b Choix des classes de recommandation

Pour notre plateforme on va utiliser deux classes de recommandation qui sont les recommandations basées sur le contenu et les recommandations de filtrage collaboratif, pour le premier jeu de données (dataset) on va l'utiliser sur les recommandations de filtrage

collaboratif et le deuxième jeu de données va être utilisé dans les recommandations basées sur le contenu.

3.6 Conclusion

Dans ce chapitre on a expliqué les différentes approches utilisées pour nos systèmes de suivi et d'assistance intelligents ainsi que leur fonctionnement et les modèles utilisées sur chaque une et leur importance dans notre plateforme, ainsi que l'importance de la génération des jeux de données à partir de notre plateforme pour être exploitée ultérieurement.

Assister l'utilisateur dans la compréhension d'un programme ou dans ses choix de formations et des challenges dans ce siècle est indispensable pour aider le client à bien apprendre et à bien s'approfondir dans le domaine ciblée.

Les moteurs de recommandation et les chatbots ont donné un très grand bénéfice aux entreprises dans ces derniers temps d'où la majorité des choix des utilisateurs on était fait à cause des moteurs de recommandation et les chatbots ont aidé des millions des utilisateurs à trouver leur réponses et trouver leur chemins.

Chapitre 4 : Conception et mise en œuvre

4.1 Introduction

Après avoir exposé des notions dans les chapitres précédents sur la sécurité informatique et ses bases et voir l'importance de la simulation et les tests d'intrusions pour renforcer la sécurité ainsi que la nécessité d'apprendre ces concepts dans notre siècle et comment une assistance intelligente peut être utile et fiable pour bien diriger l'utilisateur, pour mettre tout ça à la disposition des utilisateurs il lui faut une plateforme bien étudiée selon les demandes des clients et structurée selon son goût.

Dans ce chapitre nous allons parler de la conception de notre plateforme, ainsi que l'architecture réseau proposée pour les serveurs pour leur permettre d'héberger les différents challenges et les simulations dans des environnements sécurisé et rapide, et enfin avec une analyse des besoins et aux données utiliser pour comprendre les besoins et les demandes des utilisateurs aujourd'hui.

La phase de l'implémentation ou mise en œuvre dans ce chapitre met l'accent sur les différentes technologies utilisée pour le développement de notre plateforme avec une présentation spécifique du Framework « BotPress » qui est conçu pour la construction des chatbots, ainsi qu'une visualisation des interfaces de notre plateforme en expliquant les différents composants de cette dernière.

4.2 Analyse et conception

La phase de l'analyse et la conception est une phase très importante avant l'implémentation de la plateforme, elle permet de définir les besoins et les fonctionnalités de la plateforme, ainsi elle permet de fixer les acteurs principaux et les acteurs secondaires.

Les besoins et les fonctionnalités sont tirée à partir d'une l'analyse détaillée du cahier de charge et la description fournis et après discutée avec les différents membres du projet et de notre entreprise.

Le but de la conception est de fixer les choix techniques et de préparer l'implémentation ainsi elle doit servir de support pour l'implémentation et la maintenance, elle n'est souvent pas compréhensible par les utilisateurs mais par les développeurs.

La conception dans notre projet est l'UML. Elle s'appuie sur l'utilisation des différents diagrammes qui vont être citée et expliquée dans les sous-sections suivantes.

4.2.1 Diagrammes de cas d'utilisation

Le diagramme de cas d'utilisation nous présente les principales fonctions du système, ainsi que les acteurs qui interviennent.

Acteur : Administrateur

Il gère les fonctionnalités de la plateforme « Intervalle Security » et gère les différents challenges et les documentations et les autres composants présents dans la plateforme d'où il permet de faire :

- Mise à jour des challenges : permet l'ajout, la modification et la suppression des challenges.
- La gestion des machines : tout concernant les machines et aussi permet la mise à jour (ajout, modification, suppression).
- La gestion des catégories des challenges.

Ci-dessous dans la figure 24 une illustration du diagramme de cas d'utilisation d'un administrateur.

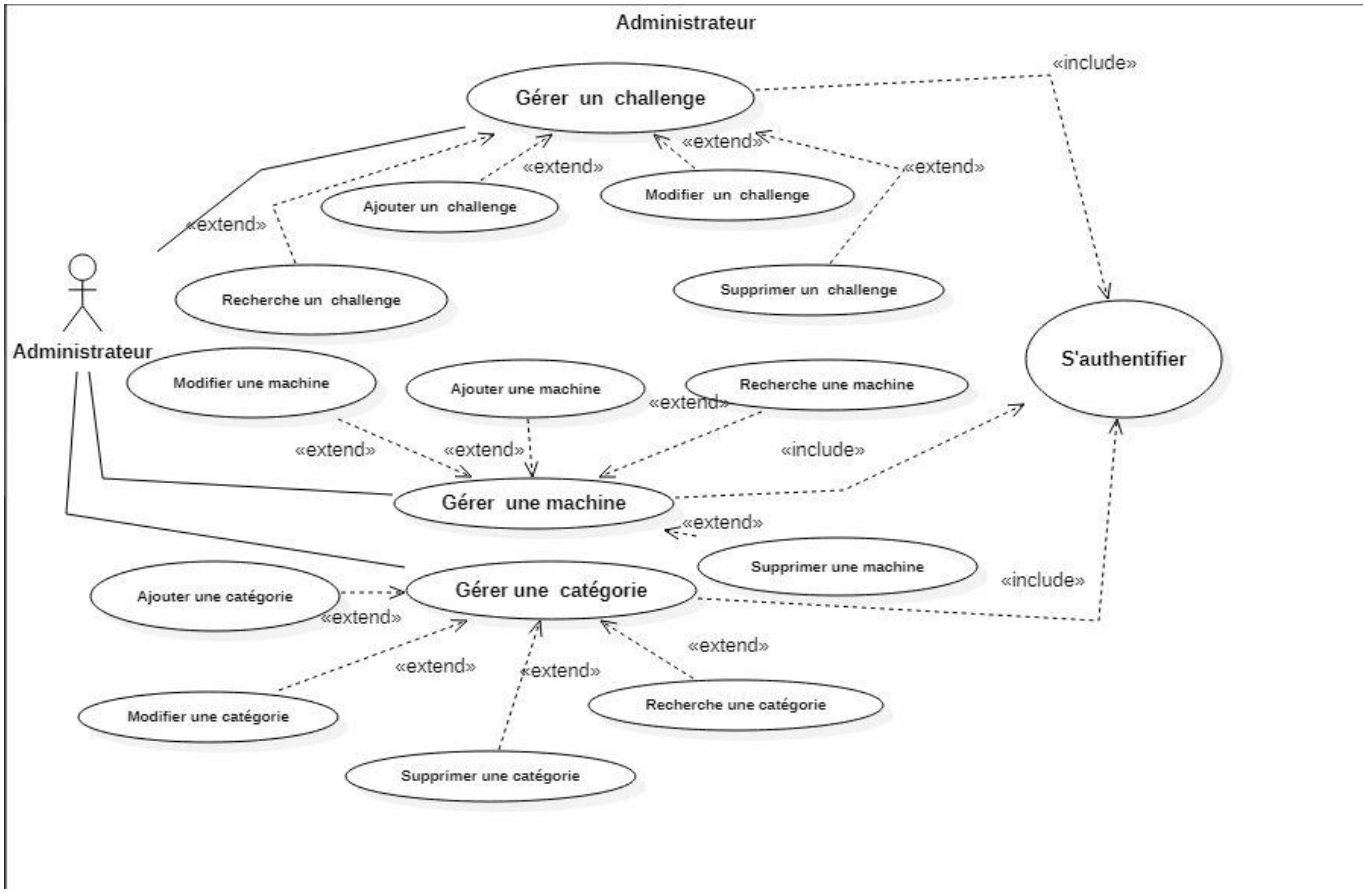

Figure 24 - Diagramme de cas d'utilisateur administrateur

Acteur : Client

Le client a plusieurs possibilités comme :

- L'accès aux challenges : l'utilisateur a la possibilité d'accéder aux challenges selon la catégorie et voir les statistiques de chaque challenge et les avis avec la possibilité bien sure de valider un flag.
- L'accès aux machines wargame : l'utilisateur a la possibilité encore d'accéder et de lancer des instances des différentes machines avec le système de validations des flags et encore il aura un certificat lors de la validation.
- La consultation des documents
- La consultation des outils
- L'accès aux simulations Red Team vs Blue Team.

Ci-dessous dans la figure 25 une illustration du diagramme de cas d'utilisation d'un administrateur.

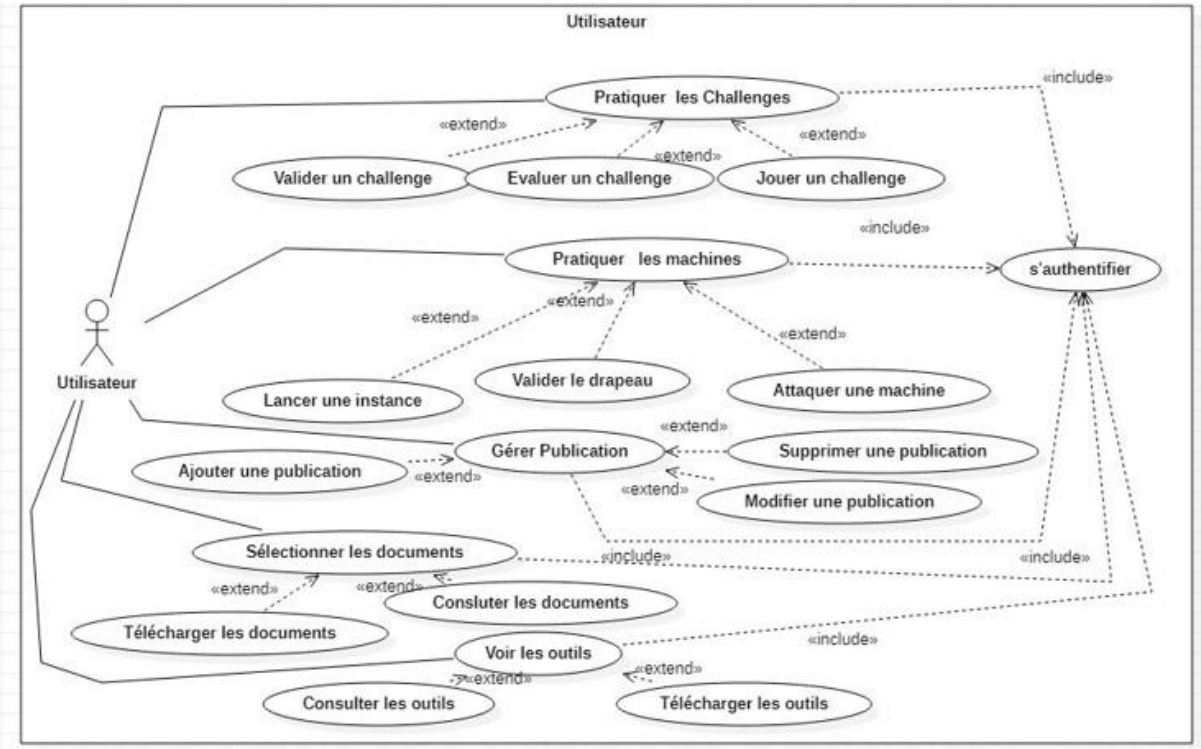

Figure 25 - Diagramme de cas d'utilisateur client

4.2.2 Diagrammes de séquences

Il permet de décrire les scénarios de chaque cas d'utilisation en mettant l'accent sur la chronologie des opérations en interaction avec les objets.

Ce diagramme met en scène une interaction. En particulier, il montre aussi les objets qui participent à cette même interaction par leur « ligne de vie » et les messages qu'ils échangent présentés sous forme de séquence dans le temps.

Ci-dessous une description des différents diagrammes de séquences de chaque cas utilisation.

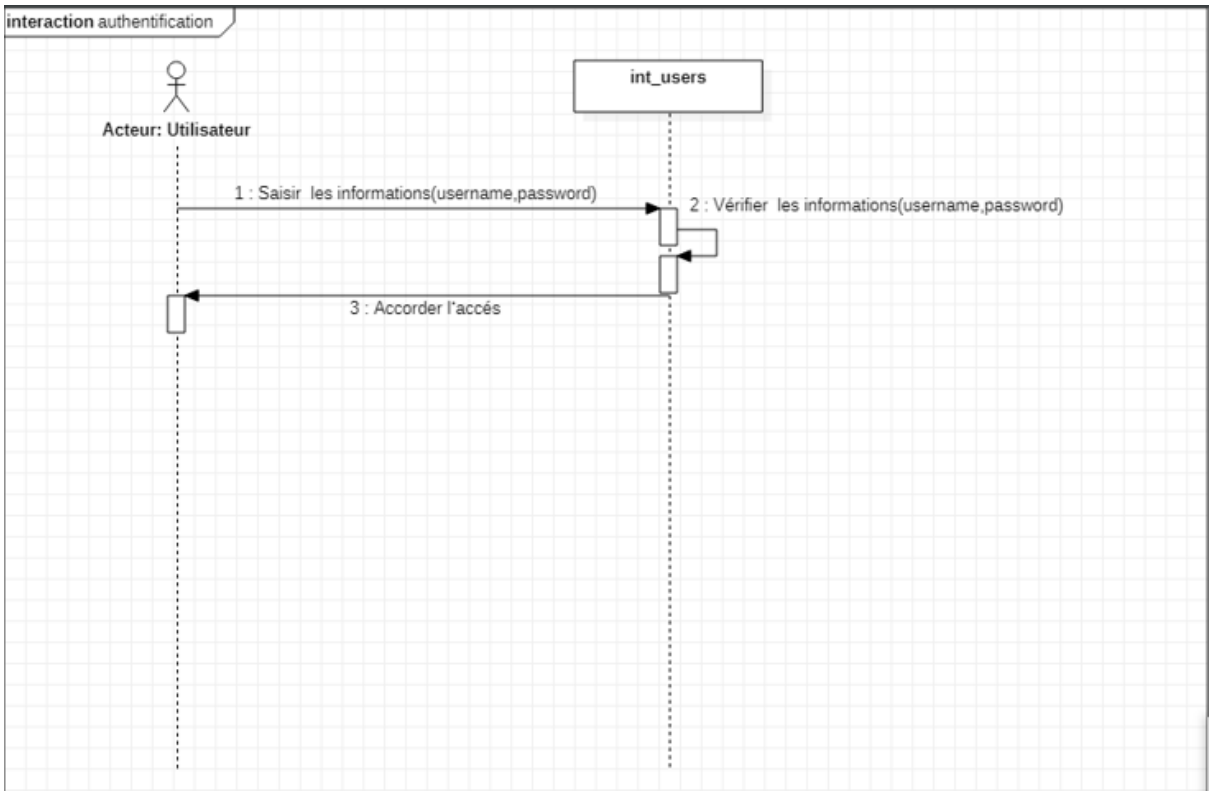

Figure 26 - Diagramme de séquence - Authentification

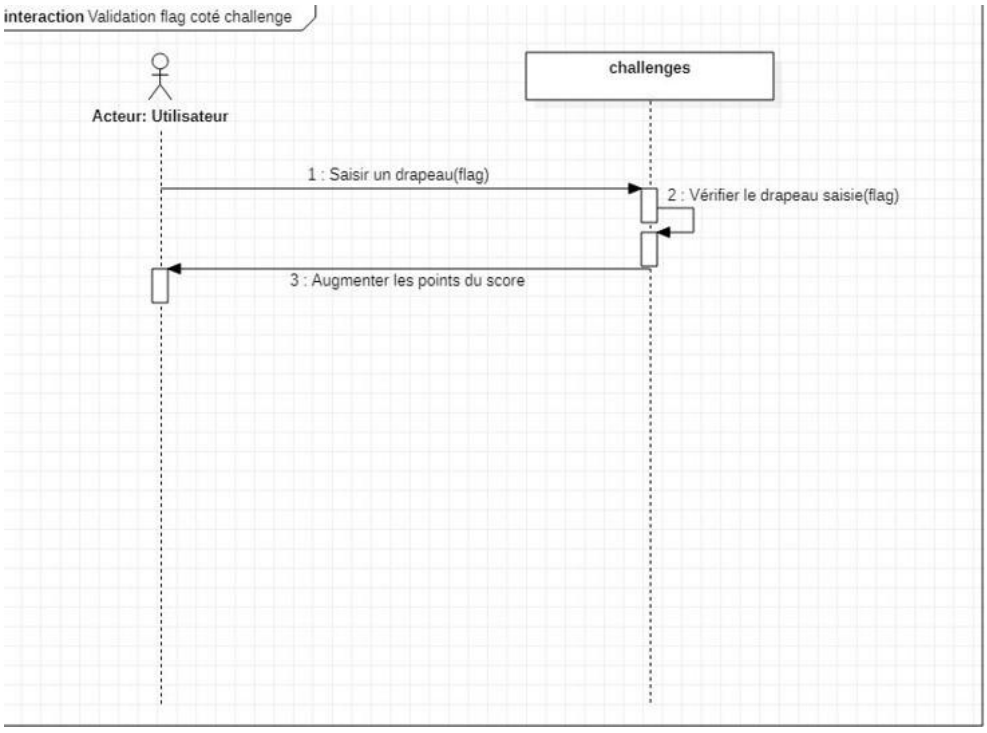

Figure 27 - Diagramme de séquence - Validation flag d'un challenge

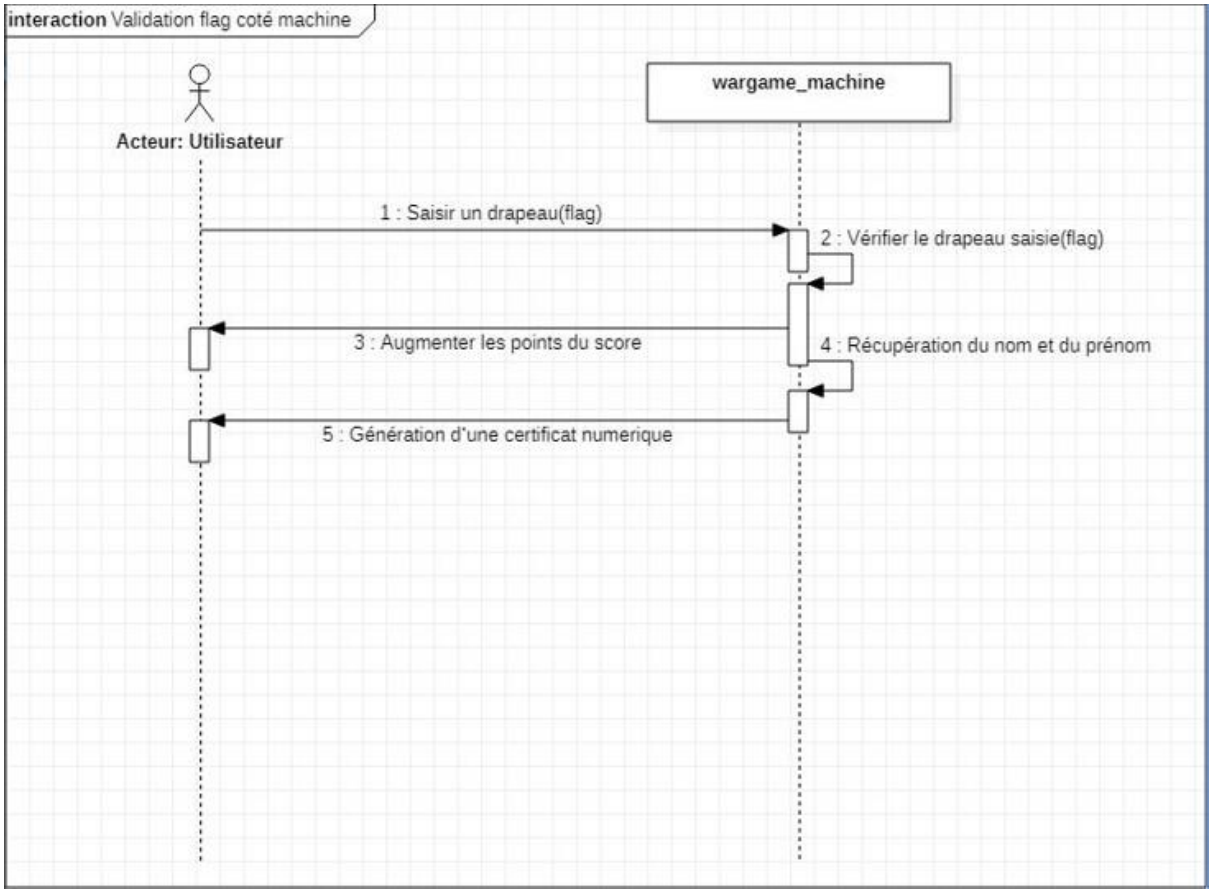

Figure 28 - Diagramme de séquence - Validation flag machine

4.3.3 Diagramme de classes

Le diagramme qui suivent **(Voir Figure 29)**, représentent les différentes classes métiers de notre plateforme. Il modélise la structure de la plateforme « Intervalle Security ».

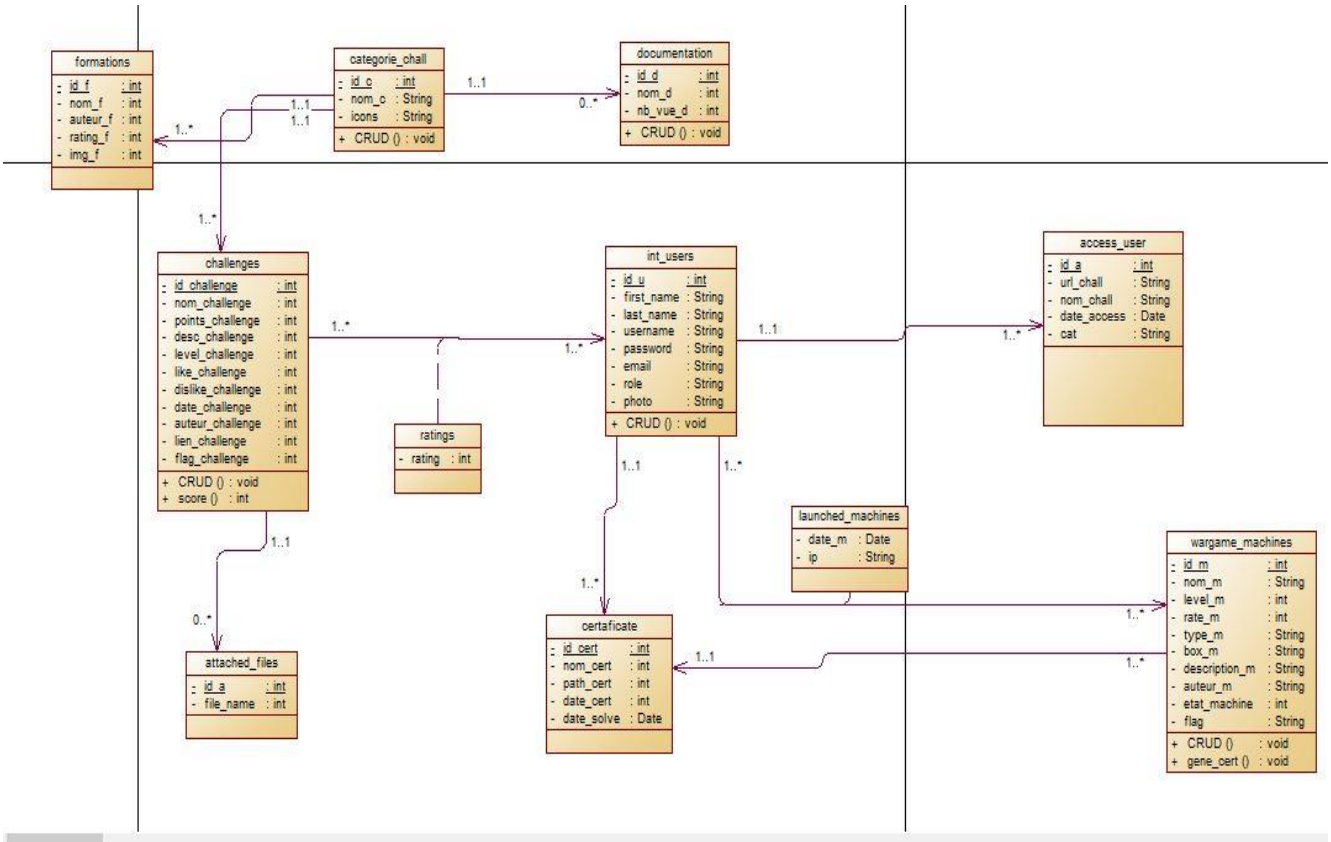

Figure 29 - Diagramme de classes

4.3 Implémentation

Après avoir vu dans la section précédente l'importance de la conception et les fonctionnalités qui peuvent êtres implémentées dans notre plateforme maintenant dans cette section c'est de voir comment implémenter et utiliser les schémas de cette conception pour développer cette plateforme, on va présenter la plateforme avec ces objectifs, ainsi que l'architecture proposée pour la plateforme et les outils et les environnements utilisée pour cette dernière.

4.3.1 Présentation de notre plateforme « Intervalle Security »

Figure 30 - Logo d'intervalle Security

Dans cette section nous allons présenter notre plateforme avec toutes les fonctionnalités, les composants qu'elle contient ainsi que le but et l'objectif de cette plateforme.

La plateforme « Intervalle Security » est une plateforme web de simulation et des tests d'intrusion dont l'objectif principale de mettre les utilisateurs dans un monde réel de sécurité et les apprendre à bien manipuler le domaine, cependant la plateforme a d'autres objectifs secondaire comme :

- Tester les compétences et les connaissances des employées de l'entreprise.
- Entraîner les employées sur les différentes parties de la sécurité informatique.
- Donner des exemples des simulations réelles aux employées.
- Améliorer les compétences des utilisateurs dans les entreprises.
- Analyse des résultats obtenus par la plateforme pour des besoins commercial ou privés, et les exploiter pour créer des systèmes intelligents.

La plateforme contient beaucoup de fonctionnalité importante qui a pour le but d'aider l'utilisateur a bien apprendre, les fonctionnalités de plateforme vont être expliquée après dans les sections qui suivent.

4.3.2 Environnement du travail

Nous allons détailler les environnements utilisés dans l'implémentation de notre plateforme

4.3.2.a Environnement matériel

Nous avons élaboré ce travail sur un PC dont la configuration est la suivante :

Pc portable:

- Marque: Lenovo.
- RAM: 8GB.
- Disque dur: 1000 GB HDD.
- Microprocesseur : Intel(R) Core(TM) i5-5200u CPU 2.2 GHz.
- Carte Graphique : AMD Radeon 8500M.

4.3.2.b Environnement Logiciel

L'environnement logiciel employé s'illustre en :

- Un système d'exploitation Kali Linux (debian 9).
- Codelite comme éditeur de texte web.
- Botpress comme Framework et éditeur dans la conception et l'intégration des chatbots.
- Jupyter Notebook pour python et l'entrainement des modèles de la machine learning ainsi que la science des données et les statistiques.
- Virtual Box pour les machines virtuelles.

4.3.3 Technologies et outils utilisés

On a parlé des environnements matériels et logiciels utilisés dans l'implémentation dans les sous-sections précédentes, maintenant on va parler des outils et des technologies utilisés pour la conception et pour le développement de notre plateforme ainsi qu'une brève présentation de chaque une de ces outils.

4.3.3.a Outils de conception

StarUML : StarUML est un logiciel open source de modélisation UML.

Power AMC : PowerAMC est un logiciel de modélisation. Il permet de modéliser les traitements informatiques et leurs bases de données associées.

4.3.3.b Outils d'implémentation

Langages de développements

Python : Python est le langage de programmation le plus utilisé dans le domaine du Big Data et du Machine Learning.

JavaScript : JavaScript est un langage de programmation qui permet d'implémenter des mécanismes complexes sur une page web, il permet de dynamiser les pages web et implémenter des évènements ou des actions sur le web.

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.

HTML : C'est un langage de balisage utilisé pour la création de pages web, permettant notamment de définir des liens hypertextes, on peut dire que c'est le langage de la structure d'une page web et il n'est pas un langage de programmation.

CSS : C'est le langage qui permet de faire un style aux pages web, notamment utilisés avec HTML.

Framework

Django : Django est un Framework de développement Web Python open source. Il vise à rendre le développement Web 2.0 simple et rapide.

BotPress : BotPress est un Framework open source conçu pour la conception et la création des chatbots, facile et rapide a utilisé.

Technologies

Apache2 : Le logiciel gratuit Apache HTTP Server est un serveur HTTP créé et maintenu dans « Apache Foundation ». En avril 2019, c'était le serveur HTTP le plus populaire sur le World Wide Web. Il est distribué sous les termes de la licence Apache

Web : Le World Wide Web s'appuie sur la notion d'architecture client-serveur. Un serveur est une machine en général assez puissante qui fournit un ou plusieurs applications web.

Bibliothèques

Toutes les bibliothèques citées ci-dessous c'est des bibliothèques python notamment utilisé dans le développement ou dans l'intelligence artificielle et de la Machine Learning.

Pandas : Pandas est une bibliothèque écrite pour le langage de programmation Python permettant la manipulation et l'analyse des données.

Scikit-learn : Scikit-learn est une bibliothèque libre Python destinée à l'apprentissage automatique.

TensorFlow : TensorFlow est un outil open source d'apprentissage automatique développé par Google. C'est l'outil le plus populaire et utilisé dans le monde.

Keras : La bibliothèque Keras permet d'interagir avec les algorithmes de réseaux de neurones profonds et d'apprentissage automatique prêt à utiliser sans besoin de les implémenter.

Endesive : Cette bibliothèque est conçue pour la signature digitale du tout type de document (PDF, XML, etc...) et aussi la vérification des signatures.

OpenSSL-pyhon : OpenSSL est une boîte à outils de chiffrement comportant deux bibliothèques, libcrypto et libssl, fournissant respectivement une implémentation des algorithmes.

Pyvbox : Pyvbox est une bibliothèque python conçue pour la manipulation des machines virtuelles instanciée par le logiciel « Virtual Box ».

4.3.4 Architecture de la plateforme

Dans la section précédente on a vu à partir de la présentation de la plateforme les différentes fonctionnalités dont les challenges, les simulations et les machine comme challenge réel, maintenant comment la plateforme interagit avec les demandes des clients, et quelle architecture va suivre pour ça et le plus important c'est quelle architecture réseau la plateforme va utiliser pour séparer et organiser les différentes parties des systèmes.

4.3.4.a Architecture Client-Serveur

L'environnement client/serveur désigne un mode de communication organisé par l'intermédiaire d'un réseau et d'une interface Web entre plusieurs ordinateurs. Cela signifie que des machines clientes (machines faisant partie du réseau) contactent un serveur, une machine généralement très puissante en termes de capacités d'entrées-sorties, qui leur fournit des services. Lequel services sont exploités par des programmes, appelés programmes clients, s'exécutant sur les machines clientes. [28]

Ci-dessous **(Voir Figure 31)** une illustration de l'architecture client-serveur comment elle fonctionne :

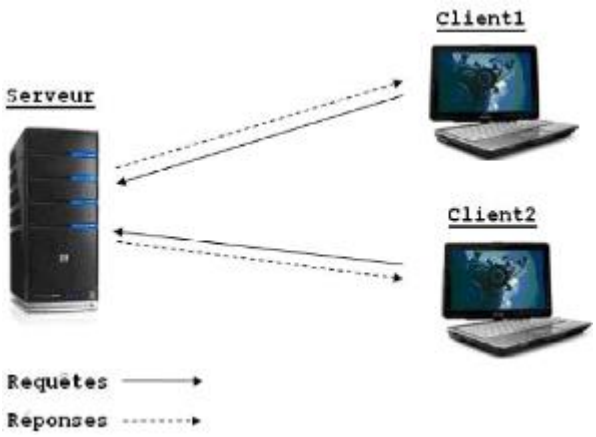

Figure 31 - Architecture client – serveur. [28]

4.3.4.b Architecture Réseau

Ci-joint **(Voir Figure 32)** l'architecture réseau proposée pour l'organisation des différentes parties de la plateforme

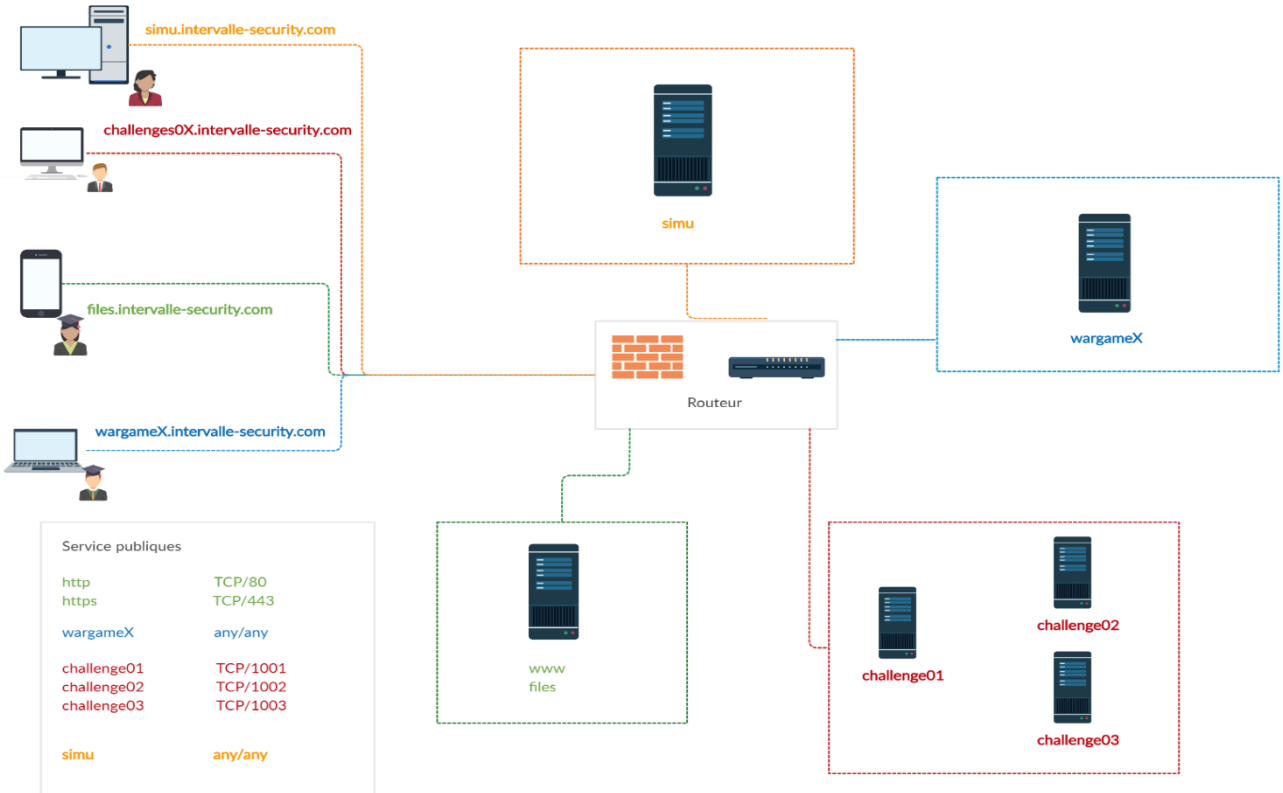

Figure 32 - Architecture réseau de la plateforme

4.3.5 Interfaces et fonctionnement de la plateforme

Dans cette section on va présenter les interfaces de la plateforme des deux côtés : coté client et coté serveur.

4.3.5.a Coté client

Dans cette partie on va présenter l'interface de notre plateforme coté utilisateur ou client avec des captures d'écran.

Login

Le coté client commence par une page de login ou authentification simple et avec des illustrations qui donne aux utilisateurs la motivation des attaques contre la défense **(Voir Figure 33)** :

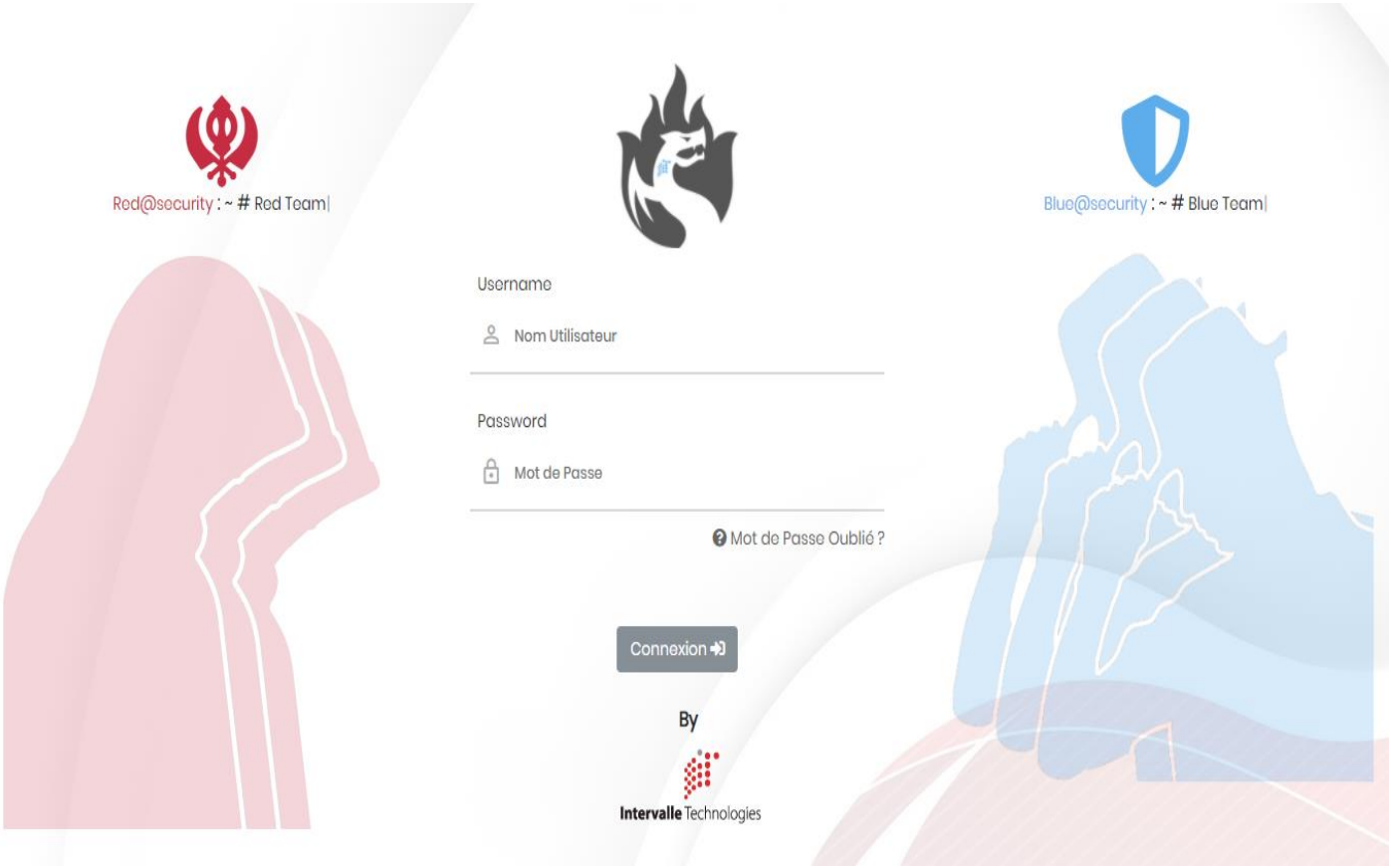

Figure 33 - Capture d'écran de la page login

Accueil

Après avoir mis un pseudo et mot de passe correct pour l'authentification dans la plateforme l'utilisateur va être redirigé dans la page d'accueil **(Voir Figure 34)**.

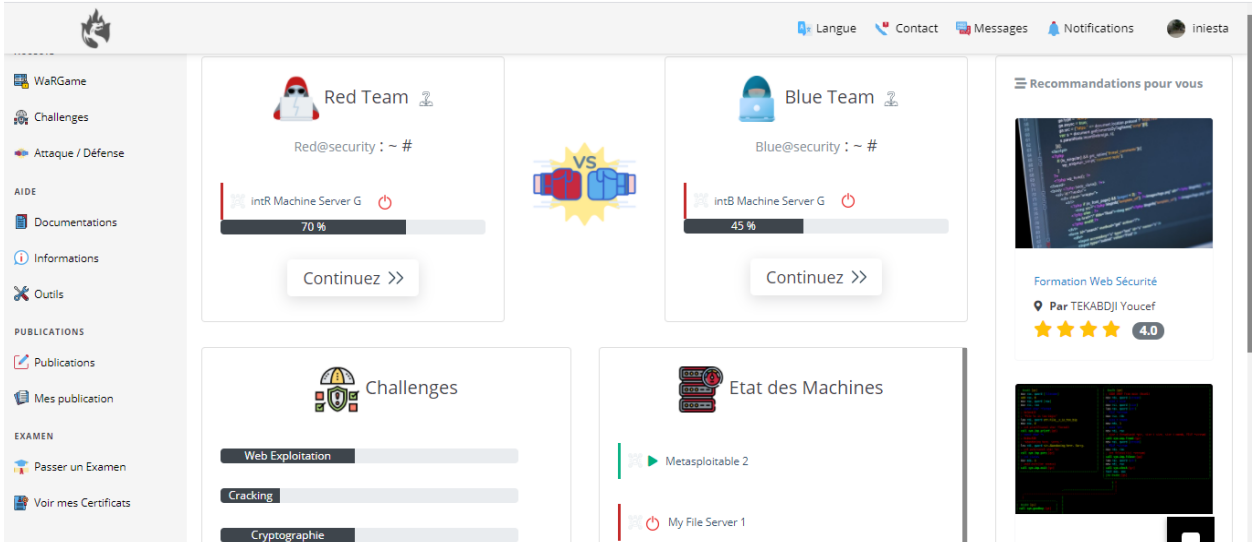

Figure 34 - Capture d'écran de la page d'accueil

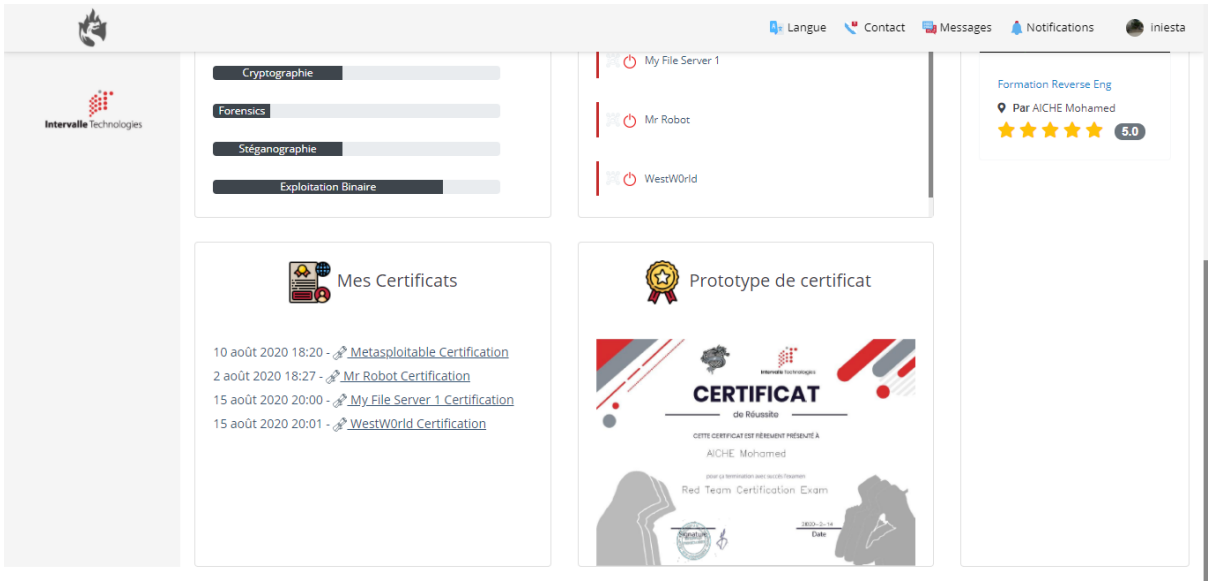

Figure 35 - Capture d'écran de la page d'accueil - 2

La partie gauche on trouve le menu qui nous mène aux différentes parties qui vont être présenté après de la plateforme avec une structure facile à utiliser.

Au centre on trouve les statistiques de l'utilisateur, l'état des systèmes et machines ainsi que les certificats obtenu par l'utilisateur.

Enfin à droite on trouve les recommandations des formations proposés par nos systèmes de recommandation intelligente pour l'utilisateur.

En coté bas à droite un bouton de chatbot est placée pour discuter avec ce dernier et être guidée par lui.

Chatbot

Comme citée précédemment en bas droite un bouton chatbot est placé pour lancer la discussion montrée ci-dessous **(Voir Figure 36)**.

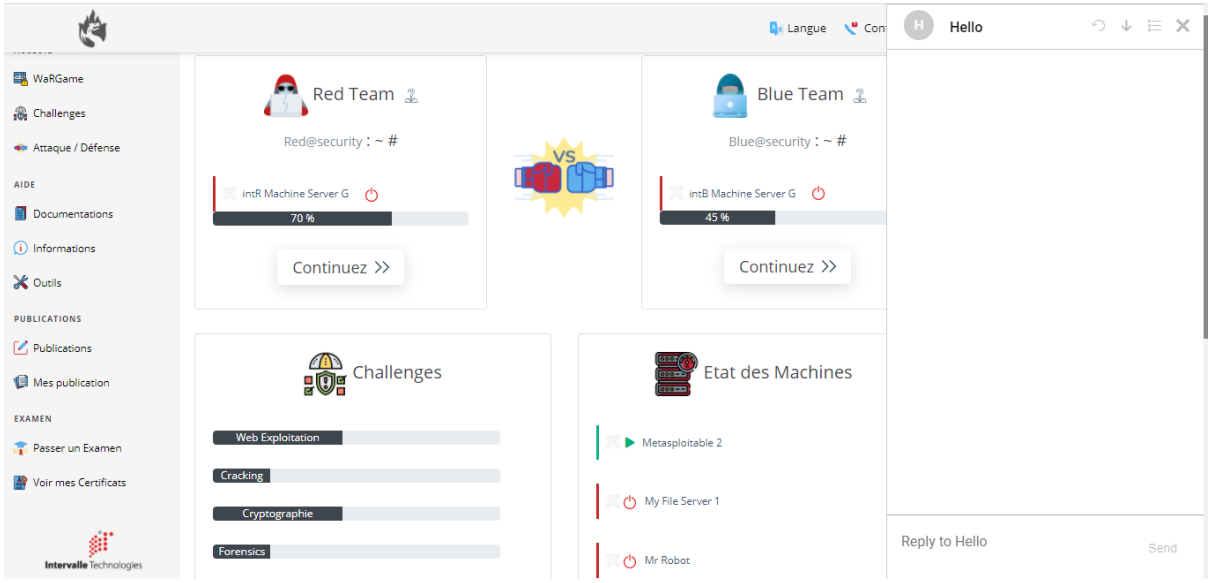

Figure 36 - Capture d'écran de la page d'accueil avec chatbot

Wargame

Maintenant pour les parties de menu à gauche, on commence par le premier « Wargame » **(Voir Figure 37)** :

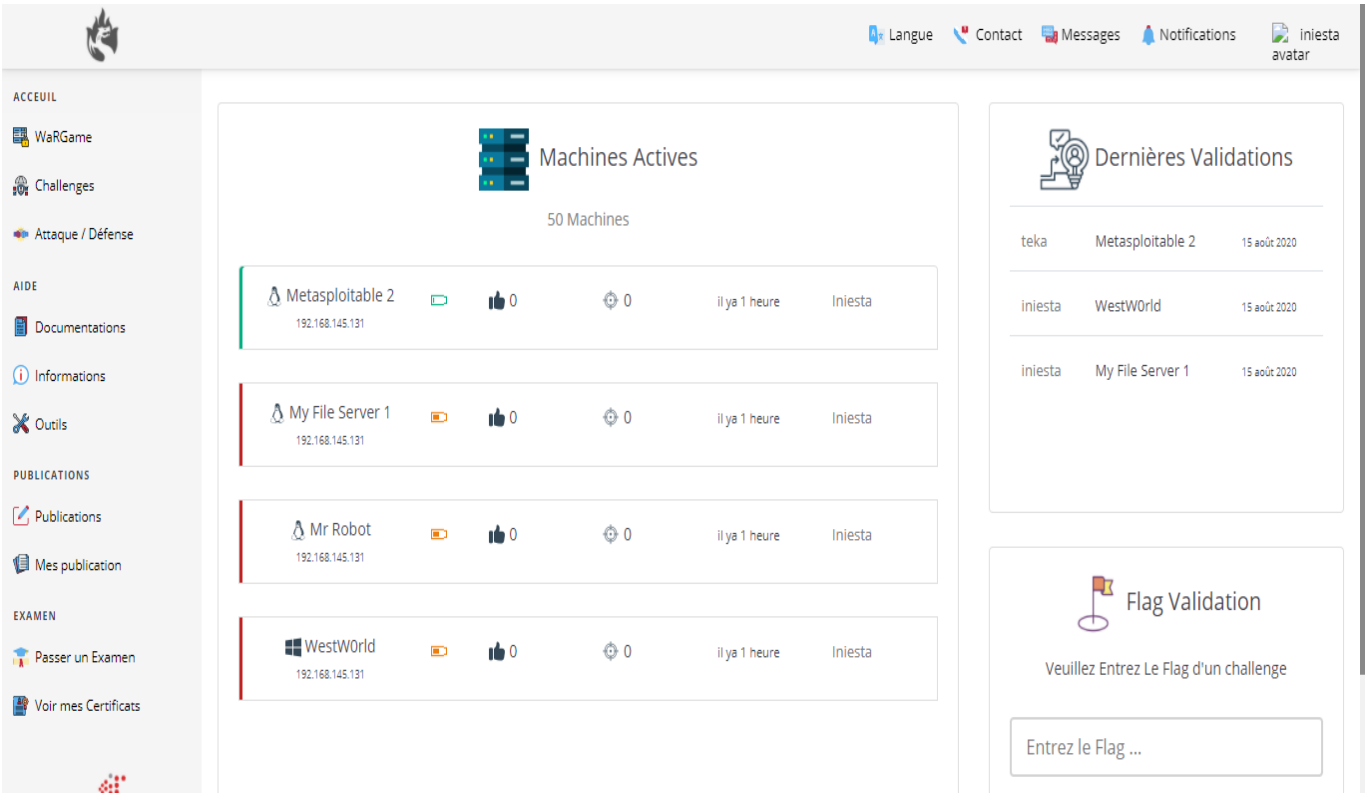

Figure 37 - Capture d'écran de la page machines

Ici on trouve les différentes machines à exploiter avec leur niveau de difficulté, les auteurs de la machines les points etc... , le couleur verte devant la machine signifie que la machine est en marche, la couleur rouge signifie qu'elle est éteinte.

A gauche on trouve les statistiques des validations, et la validation des flags, en validant un flag une certificat sous forme PDF avec une signature numérique va être offerte pour chaque machine résolue **(Voir Figure 38)**.

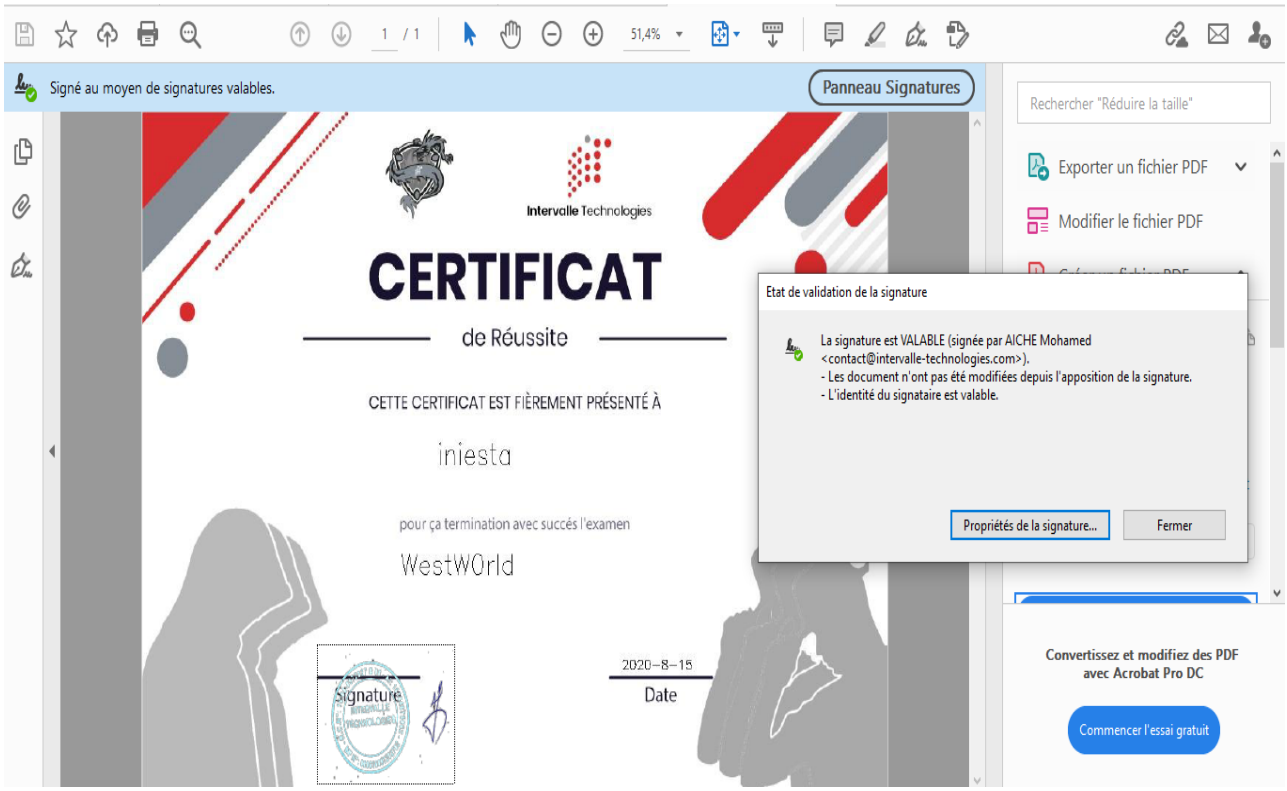

Figure 38 - Exemple d'une certificat numérique

Challenges

Maintenant pour la partie intéressante de la plateforme, c'est la partie des challenges et des tests d'intrusion.

Chaque challenge a un titre et le nom d'auteur qui peuvent aider à résoudre les challenges ainsi que le niveau de difficulté, les avis des utilisateurs etc…

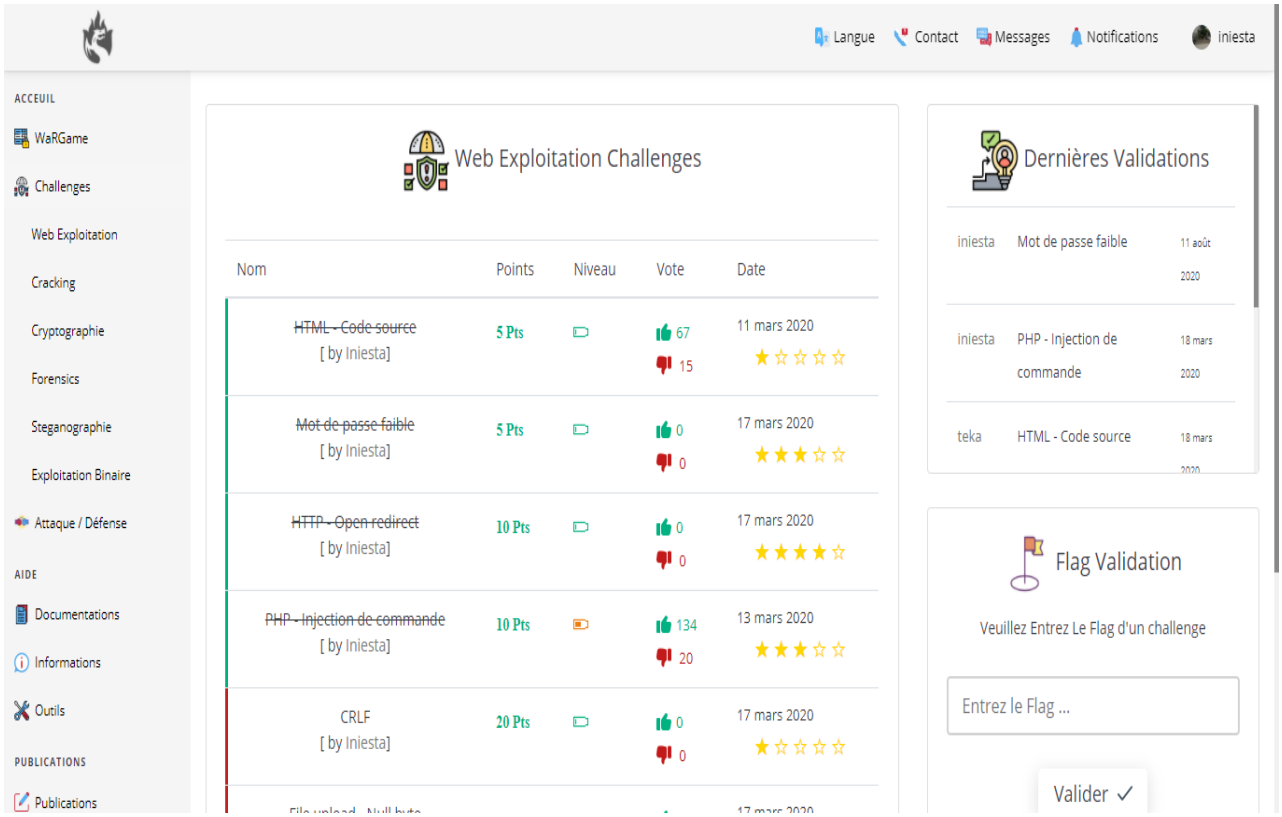

Figure 39 - Capture d'écran de la page des challenges

Les challenges résolus sont avec un trait vert, et aussi il y a le système de validation des flags.

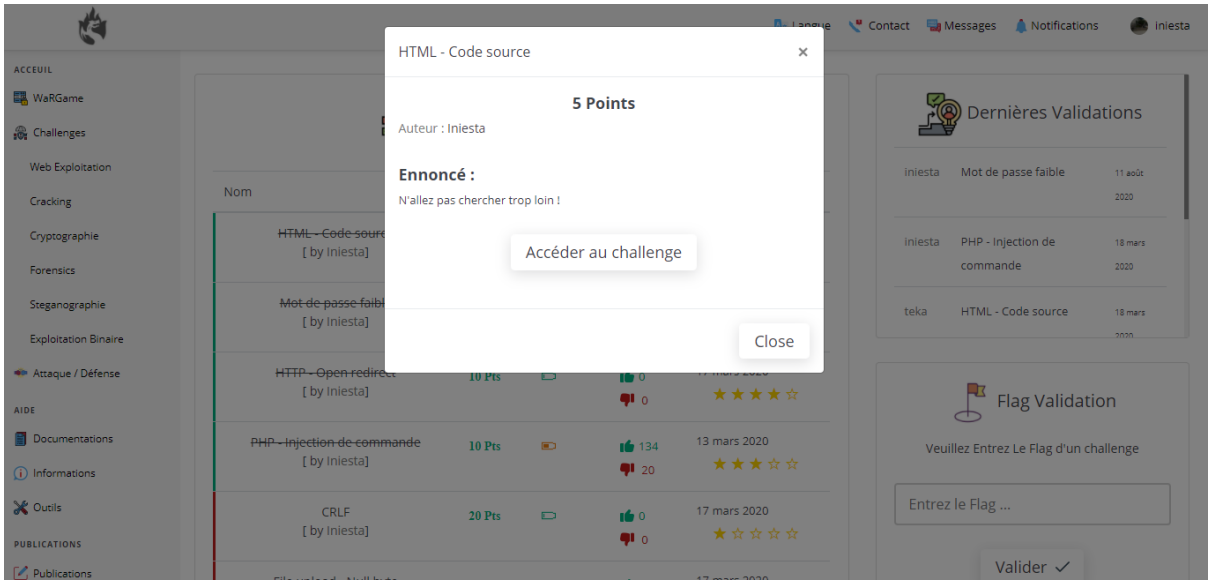

Figure 40 - Capture d'écran challenge

Un énoncé spécifique au challenge est donné pour aider à la compréhension et la résolution du challenge.

Documentation

La documentation est importante pour les débutants dans le domaine pour avoir une bonne compréhension des challenges.

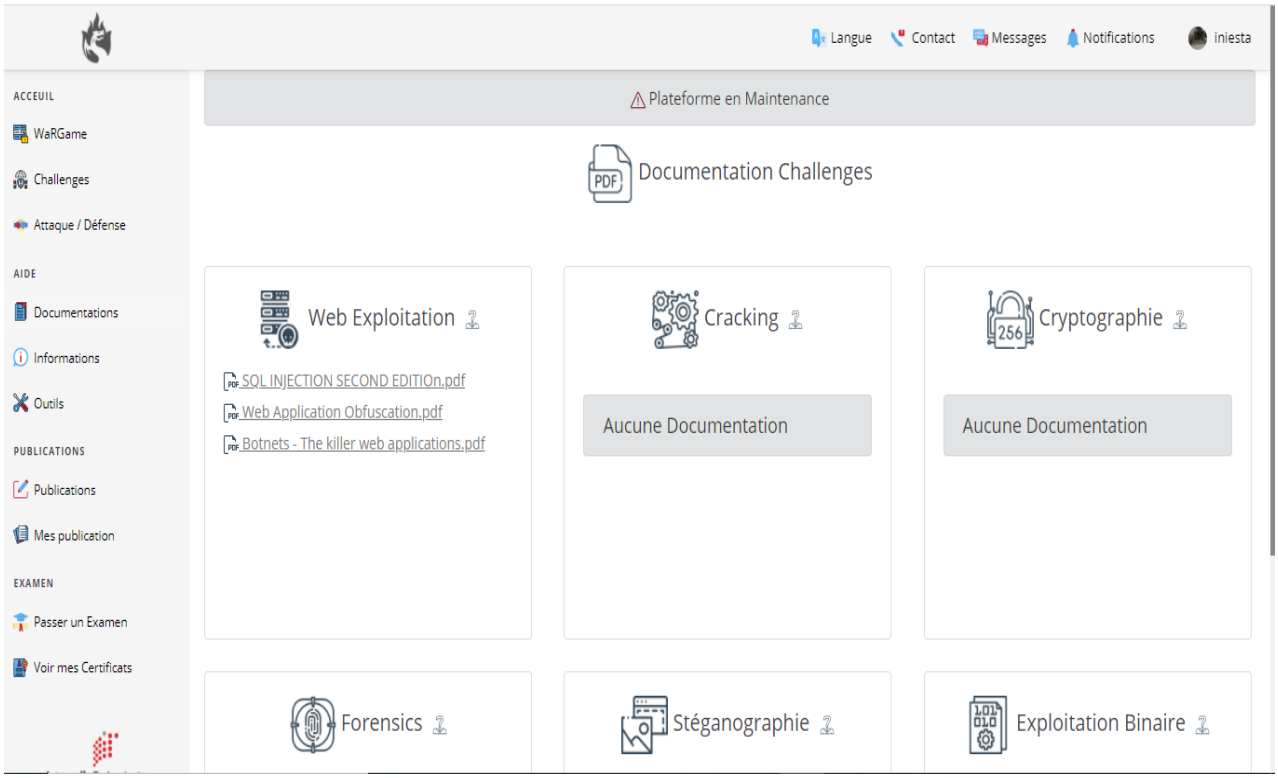

Figure 41 - Capture d'écran de la page documentation

Ces parties sont les parties essentielles dans la plateforme coté client, maintenant dans la sous-section suivante on va aborder le coté admin.

4.3.5.b Coté admin

La création des challenges, la configuration des machines et tout ce qui concerne la gestion, cette partie du coté admin est une partie consacré juste pour les administrateurs de la plateforme pour gérer la plateforme.

Gestion des challenges

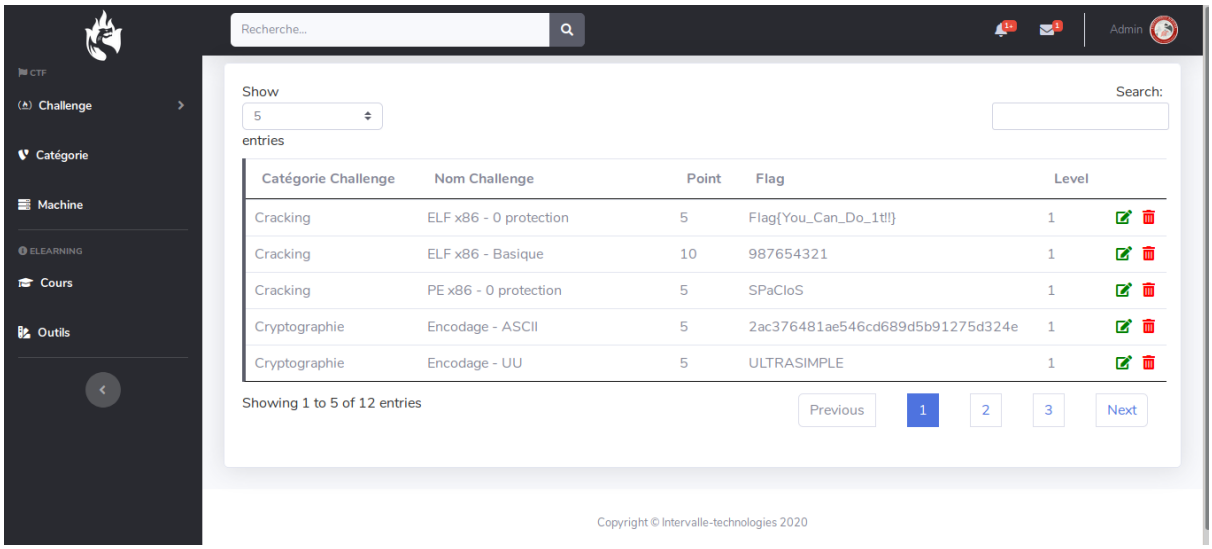

Figure 42 - Capture d'écran de la gestion des challenges

Dans la partie c'est dessous, nous montrons la page de la configuration des challenges, leur niveau, flags ainsi que leur nom et autres.

Gestion des machines

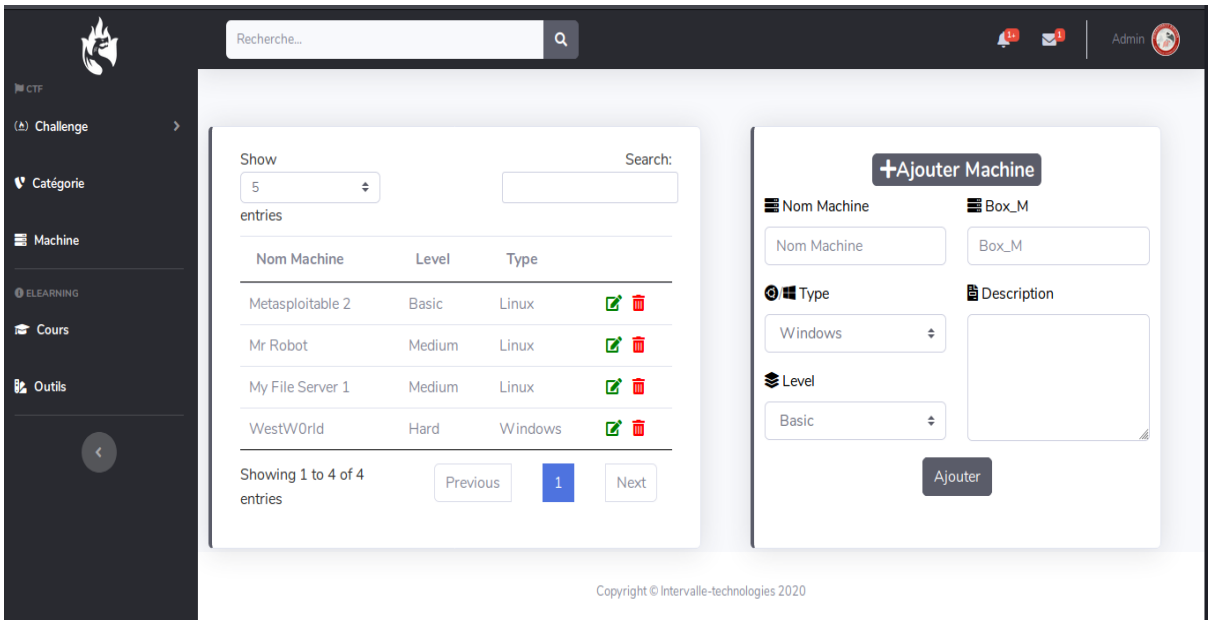

Figure 43 - Capture d'écran de la page des machines

Configuration des catégories des challenges

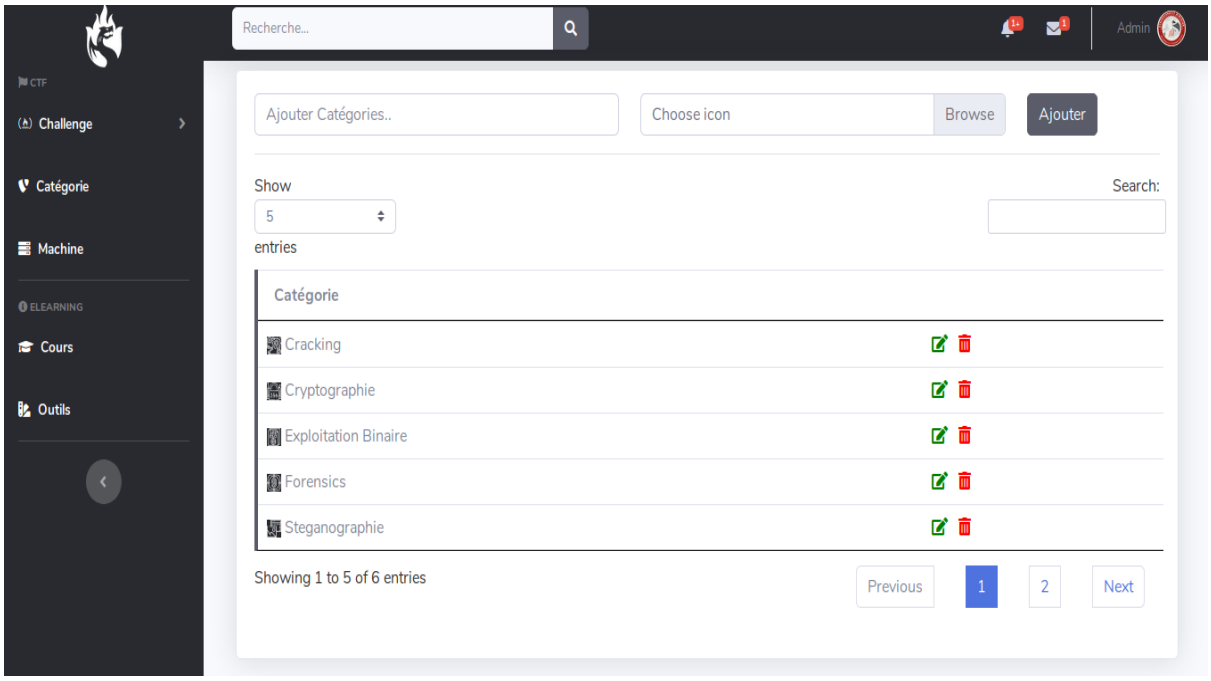

Figure 44 - Capture d'écran de la page des catégories

4.4 Installation

Dans cette étape on va décrire comment installer la plateforme « Intervalle Security » dans un système.

Python3 est nécessaire pour l'installation de la plateforme, donc au début il faut installer python3.

Après ouvrir le « terminal » dans linux ou « cmd » et exécuter les commandes suivantes pour installer les bibliothèques nécessaires pour le bon fonctionnement de la plateforme :

- # pip3 install endesive
- # pip3 install pyvbox
- # pip3 install django
- # pip3 install pandas
- # pip3 install sckit-learn
- # pip3 install keras
- # pip3 install python-opencv

Après l'exécution de ces commandes et aucun problème n'est survenu.

Nous allons va tester le lancement de la plateforme avec la commande suivante :

- # cd intervalle
- # python3 manage.py runserver <ip>:<port> (l'ip et le port sont choisis par l'utilisateur) **(Voir Figure 45)**

```
root@Iniesta:~/Bureau# ls
intervalle  rockyou.txt
root@Iniesta:~/Bureau# cd intervalle/
root@Iniesta:~/Bureau/intervalle# ls
cachet.png      chall       db.sqlite3  exams       intervalle.p12  test.jpeg
certificat.jpg  challenges  demo.p12    intervalle  manage.py       wargame
root@Iniesta:~/Bureau/intervalle# ls
cachet.png      chall       db.sqlite3  exams       intervalle.p12  test.jpeg
certificat.jpg  challenges  demo.p12    intervalle  manage.py       wargame
root@Iniesta:~/Bureau/intervalle# python3 manage.py runserver 0.0.0.0:8000
Watching for file changes with StatReloader
Performing system checks...

System check identified no issues (0 silenced).
August 21, 2020 - 22:00:35
Django version 3.0.8, using settings 'intervalle.settings'
Starting development server at http://0.0.0.0:8000/
Quit the server with CONTROL-C.
```

Figure 45 - Lancement de la plateforme

4.5 Conclusion

Ce chapitre nous a donné une vision générale sur notre plateforme et il a donné l'aspect conceptuel de la plateforme « Intervalle Security » à travers les différents diagrammes décrits en UML, des diagrammes qui illustre les différents cas d'utilisation de la plateforme, et des diagrammes de séquences qui illustre le processus de visualisation de la plateforme, et des diagrammes de classes, ainsi que des besoins et la présentation de la plateforme.

La vision conceptuelle de la plateforme nous a permis d'avoir une idée générale sur le processus du développement et l'implémentation, ainsi ce chapitre nous a présenter la plateforme et ces fonctionnalités avec les différents outils et technologies utilisés pour implémenter cela dans un environnement matériel et logiciel expliquée, l'illustration des interfaces la plateforme donne une bonne vision pour connaitre les fonctionnalités et les capacités de la plateforme.

Conclusion générale

Dans ce mémoire nous nous sommes intéressés dans la lutte contre les cyberattaques en intégrant l'apprentissage par la pratique dont on utilise le concept des tests d'intrusions et de la simulation sur les côtés défenses et attaques, ainsi qu'une assistance intelligente pour guider cet apprentissage.

Après l'analyse des statistiques des autres plateformes d'apprentissage on constate qu'une grande variété d'utilisateurs sur les choix des domaines de sécurité avec une grande demande des utilisateurs d'une assistance et des ressources qui peuvent aider ce dernier a bien cibler son objectif.

Notre plan de l'implémentation de plateforme était de se concentrer sur les besoins des utilisateurs pour apprendre en analysant leur interactions et leur demande sur nos systèmes, en analysant par exemple l'utilisateur combien de challenge d'une catégorie spécifique à accéder ou aux avis des utilisateurs sur une catégorie spécifique peut très bien aider les administrateurs de la plateforme à comprendre les besoins.

Les statistiques et jeu de données dans le domaine de la sécurité informatique et surtout pour les moteurs de recommandations et les chatbots sont presque indisponibles aujourd'hui dans le monde informatique d'où à cause de ce problème nous avons rencontré trop de difficulté à gérer et à entrainer nos systèmes, ce qui nous a fait penser à créer nos propres jeux de données mais qui peuvent se limiter si l'utilisateur n'interagit pas bien avec notre plateforme.

A part l'idée de l'assistance intelligente pour aider les utilisateurs à apprendre une idée de prédire les futurs résultats des utilisateurs à partir des données du passé en utilisant le concept de Machine Learning de « Times Series Analysis » peut donner encore un avancement sur le concept de l'apprentissage.

Webographie

[1] « Principe de Sécurité Informatique **»** [En Ligne] Consulté en ligne le 15 juillet 2020, Disponible à l'adresse : https://www.cours-gratuit.com/cours-reseau/cours-sur-les-principes-de-securite-informatique

[2] « Critère fondamentaux de sécurité » [En Ligne] Consulté en ligne le 15 juillet 2020, Disponible à l'adresse : http://info-attitude.com/4-criteres-fondamentaux-securite-information/

[3] « Way to Learn X - Différence entre attaque active et attaque passive » [En Ligne] Consulté en ligne le 15 juillet 2020, Disponible à l'adresse : https://waytolearnx.com/2018/07/difference-entre-attaque-active-et-attaque-passive.html

[4] « Cours d'introduction à la sécurité des systèmes informatique » [En Ligne] Consulté en ligne le 15 juillet 2020, Disponible à l'adresse : https://www.cours-gratuit.com/cours-divers/cours-d-introduction-a-la-securite-des-systemes-informatique

[5] « Systèmes de détection d'intrusion » [En Ligne] Consulté en ligne le 16 juillet 2020, Disponible à l'adresse : https://web.maths.unsw.edu.au/~lafaye/CCM/detection/ids.htm

[6] « Qu'est-ce qu'un VPN (Virtual Private Network) et à quoi sert-il ? » [En Ligne] Consulté en ligne le 16 juillet 2020, Disponible à l'adresse : https://openclassrooms.com/fr/courses/2939006-protegez-l-ensemble-de-vos-communications-sur-internet/2940511-quest-ce-quun-vpn-virtual-private-network-et-a-quoi-sert-il

[7] « Introduction au Reverse Engineering » [En Ligne] Consulté en ligne le 19 juillet 2020. Disponible à l'adresse : http://igm.univ-mlv.fr/~dr/XPOSE2013/reverse_engineering/generality.html

[8] « Binary Exploitation » [En Ligne] Consulté en ligne le 19 juillet 2020, Disponible à l'adresse : https://ctf101.org/binary-exploitation/overview/

[9] "How well do you know Digital Forensics?" [En Ligne] Consulté en ligne le 19 juillet 2020, Disponible à l'adresse: https://www.eccouncil.org/what-is-digital-forensics/

[10] « Wikipédia » [En Ligne] Consulté en ligne le 21 juillet 2020, Disponible à l'adresse : https://fr.wikipedia.org/wiki/Signature_num%C3%A9rique

[11] « Test d'intrusion » [En Ligne] Consulté en ligne le 23 juillet 2020, Disponible à l'adresse : https://fr.wikipedia.org/wiki/Test_d%27intrusion

[12] « Simulation informatique » [En Ligne] Consulté en ligne le 23 juillet 2020, Disponible à l'adresse : https://fr.wikipedia.org/wiki/Simulation_informatique

[13] « Simulation de cyber-attaques, Européens et Américains collaborent » [En Ligne] Consulté en ligne le 23 juillet 2020, Disponible à l'adresse : https://www.lemondeinformatique.fr/actualites/lire-simulation-de-cyber-attaques-europeens-et-americains-collaborent-42495.html

[14] « La simulation d'attaques informatiques, exercice clé du ministère des Armées en matière de cyberdéfense » [En Ligne] Consulté en ligne le 23 juillet 2020, Disponible à l'adresse : https://www.usine-digitale.fr/article/la-simulation-d-attaques-informatiques-exercice-clef-du-ministere-des-armees-en-matiere-de-cyberdefense.N862150

[15] « Qu'est-ce qu'une mission Red Team ? » [En Ligne] Consulté en ligne le 23 juillet 2020, Disponible à l'adresse : https://www.hubone.fr/oneblog/cybersecurite-quest-ce-quune-mission-red-team/#:~:text=Les%20tests%20d'intrusion%20Red,soient%20techniques%2C%20physiques%20ou%20humains

[16] « Qu'est-ce qu'un service cloud ? » [En Ligne] Consulté en ligne le 21 juillet 2020, Disponible à l'adresse : https://www.citrix.com/fr-fr/glossary/what-is-a-cloud-service.html

[17] « Qu'est-ce qu'une machine virtuelle ? » [En Ligne] Consulté en ligne le 21 juillet 2020, Disponible à l'adresse : https://azure.microsoft.com/fr-fr/overview/what-is-a-virtual-machine/

[18] « Le Machine Learning au service des outils HTTPCS » [En Ligne] Consulté en ligne le 25 juillet 2020, Disponible à l'adresse : https://www.httpcs.com/fr/machine-learning-cybersecurite

[19] « Apprentissage automatique et deep learning » [En Ligne] Consulté en ligne le 25 juillet 2020, Disponible à l'adresse : https://www.stemmer-imaging.com/fr-ch/conseil-techniqu/apprentissage-automatique-et-apprentissage-profond/

[20] « Comprendre les réseaux de neurones » [En Ligne] Consulté en ligne le 26 juillet 2020, Disponible à l'adresse : https://moncoachdata.com/blog/comprendre-les-reseaux-de-neurones/

[21] « RNN or Recurrent Neural Network for Noobs » [En Ligne] Consulté en ligne le 26 juillet 2020, Disponible à l'adresse : https://hackernoon.com/rnn-or-recurrent-neural-network-for-noobs-a9afbb00e860

[22] « A Comprehensive Guide to Convolutional Neural Networks » [En Ligne] Consulté en ligne le 26 juillet 2020, Disponible à l'adresse : https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[23] « Apprentissage Supervisé Vs. Non Supervisé » [En Ligne] Consulté en ligne le 26 juillet 2020, Disponible à l'adresse : https://le-datascientist.fr/apprentissage-supervise-vs-non-supervise

[24] « Création d'un chatbot pour les débutant » [En Ligne] Consulté en ligne le 2 aout 2020, Disponible à l'adresse : https://core.ac.uk/reader/286406000

[25] « Intelligence Artificielle : quelle différence entre NLP et NLU ? » [En Ligne] Consulté en ligne le 2 aout 2020, Disponible à l'adresse : https://www.lemagit.fr/conseil/Intelligence-Artificielle-quelle-difference-entre-NLP-et-NLU

[26] « Natural Language Processing Chatbots: The Layman's Guide » [En Ligne] Consulté en ligne le 2 aout 2020, Disponible à l'adresse : https://landbot.io/blog/natural-language-processing-chatbot/

[27] « Recommender System » [En Ligne] Consulté en ligne le 16 aout 2020, Disponible à l'adresse : https://developers.google.com/machine-learning/recommendation/

[28] « Recommandation System » [En Ligne] Consulté en ligne le 16 aout 2020, Disponible à l'adresse : https://www.yash.com/blog/recommendation-system/