République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Mémoire de Projet de Fin d'Etude Licence

Faculté d'Informatique

Département IA & SD

Option : Informatique Générale

Conception et mise en oeuvre d'une solution de surveillance réseau pour la cyber-sécurité

Encadré par :
GUEMRAOUI Lila
ZEMMACHE Amina
Membre du Jurys :
ZAFOUNE Youcef
BELLACHE Mohemed

Présenté par :
BENNOUR Salah Eddine Hamza
SMAILI Mohamed Abdeldjalil

Projet : ACAD_001 / 2023

REMERCIEMENTS

En préambule à ce mémoire, nous remercions ALLAH de nous avoir aidé et donné la patience et le courage durant ces longues années d'étude.

Nous tenons à exprimer notre profonde gratitude envers nos encadrantes, Madame GUEMRAOUI Lila et Madame ZEMMACHE Amina , qui nous ont accompagnés tout au long de notre parcours pour la réalisation de notre premier mémoire. Leur précieux temps et leur assistance ont été essentiels pour mener à bien notre projet. Elles ont toujours été disponibles pour nous offrir leurs meilleurs conseils et leur expertise. Nous nous sentons extrêmement privilégiés d'avoir eu l'opportunité de travailler avec elles.

Nous tenons également à remercier Madame BENSIMESSAOUD Sihem et Madame DJELLALBIA Amina pour leur assistance et leurs précieux conseils. Nous souhaitons également exprimer notre reconnaissance envers toute l'équipe pédagogique de l'USTHB, et en particulier envers les enseignants qui s'investissent pleinement pour former les informaticiens de demain. Nous souhaitons adresser notre gratitude spéciale aux membres du jury pour l'honneur qu'ils nous ont accordé en acceptant d'évaluer ce travail.

Nous voulons également exprimer nos sincères remerciements à nos familles ainsi qu'à toutes les personnes qui ont contribué de près ou de loin à notre formation en licence. Enfin, nous remercions toutes les personnes qui ont pris le temps de lire et de s'intéresser à notre travail.

DÉDICACES

Je dédie ce travail :

A mes chers parents, qui m'ont encouragé et veillé à ce que je réussisse dans mes études.

A ma soeur et mon petit frère, ma famille, mes amis Adnane, Ouassim, Rami, Mina, Aya, Ahlem, Nada et tous ceux qui m'ont soutenu durant toutes les épreuves passées.

A mon binôme Hamza qui s'est donné à fond et qui n'a rien lâché depuis le début.

SMAILI Mohamed Abdeldjalil

À toute personne ayant cru en moi, merci.

BENNOUR Salah Eddine Hamza

RÉSUMÉ

L'évolution rapide de la technologie et l'omniprésence des réseaux informatiques ont ouvert la porte à de nouvelles formes d'attaques et de menaces qui peuvent causer des perturbations majeures et des pertes financières considérables. Dans ce contexte, la détection précoce et précise des attaques réseaux est devenue une priorité absolue pour assurer la sécurité des systèmes informatiques et protéger les données sensibles. Les approches traditionnelles de détection d'attaques, basées sur des règles préétablies et des signatures connues, se révèlent souvent insuffisantes face à la complexité et à la variété des attaques modernes. C'est là qu'intervient l'apprentissage automatique (machine learning) qui offre la possibilité d'exploiter l'énorme quantité de données générées par les réseaux pour identifier des modèles, des anomalies et des comportements suspects, permettant ainsi de détecter les attaques réseaux de manière plus efficace. L'objectif principal de notre travail est donc de proposer une solution de surveillance continue (monitoring) du trafic réseau pour la détection des attaques basée sur les techniques de l'apprentissage automatique.

Mots clés : Attaques réseaux, Apprentissage automatique, Surveillance réseau, Détection des attaques.

Abstract

The rapid evolution of technology and the omnipresence of computer networks have opened the door to new forms of attacks and threats that can cause major disruptions and significant financial losses. In this context, early and accurate detection of network attacks has become an absolute priority to ensure the security of computer systems and protect sensitive data. Traditional attack detection approaches, based on predefined rules and signatures, often prove insufficient to face the complexity and diversity of modern attacks. This is where machine learning comes into play, offering the opportunity to analyze the vast amount of network traffic to identify anomalies and suspicious behaviors, thereby enabling more effective detection of network attacks. The main objective of our work is to propose a continuous network traffic monitoring solution for attack detection based on machine learning techniques.

Keywords : Network attack, Machine learning, Network monitoring, Attacks detection.

TABLE DES MATIÈRES

TABLE DES FIGURES

LISTE DES TABLEAUX

Introduction générale

Dans le monde d'aujourd'hui, le numérique a investi tous les secteurs : économie, santé, télécommunications, ... etc. Tout le monde passe au numérique, parce qu'ils veulent suivre le rythme de la société et de la technologie. Le cyberespace est devenu un nouvel univers sans frontières dans lequel tous les acteurs partagent de l'information et communiquent dans tous les domaines.

Toutefois, cette dépendance vis-à-vis du numérique est devenue une arme à double tranchant. En effet, à mesure que le monde est devenu de plus en plus intégré numériquement, les cyber-risques sont devenus de plus en plus diversifiés, complexes et difficiles à contrer : Toute faille de sécurité peut être actuellement exploitée de manière malveillante. Ceci a conduit les organismes à accorder une attention renouvelée aux moyens et méthodes de la cyber-sécurité en adaptant des approches préventives qui consistent à rendre le cyberespace moins favorable à l'expression de la criminalité et à réduire les opportunités criminelles.

Ainsi, contrairement aux évaluations traditionnelles de la cyber-sécurité qui sont effectuées périodiquement, il faut analyser le cyberespace continuellement, afin de détecter des indicateurs d'attaques potentielles, d'identifier leurs types, leurs fréquences et leur sévérité, et ceci en temps opportun. La surveillance continue du cyberespace est une approche par laquelle une organisation surveille en permanence ses systèmes et réseaux informatiques pour détecter de manière automatisée les menaces de sécurité avant qu'elles ne causent des dommages et des perturbations.

Cela peut être fait à l'aide d'une variété d'outils, y compris les outils de détection d'intrusion (IDS), pare-feu, ... etc. Cependant, dans un environnement où les menaces sont de plus en plus complexes et où la quantité de données à traiter est croissante, les outils traditionnels de détection d'intrusion soulignent leurs propres limites, et s'avèrent insuffisants. Pour pallier ses limites, plusieurs applications de cyber-sécurité ont été développées, notamment celles basées principalement sur des techniques de l'apprentissage automatique (ou machine learning en anglais) qui changent la donne en matière de cyber-sécurité, et ceci en analysant des quantités massives de données sur les risques afin d'accélérer les temps de réponse et d'aider les opérations de sécurité à garder une longueur d'avance sur les menaces.

Effectivement, les données constituent la clé d'un environnement cyber optimal. L'apprentissage automatique a le potentiel d'identifier les données appropriées pour obtenir les meilleurs résultats. Ces données produites permettent de mieux appréhender les cyber-menaces qui se

profilent. C'est dans cette optique que s'inscrit notre contribution, qui vise la proposition et le développement d'une solution de surveillance continue (monitoring) du trafic basée sur les techniques de l'apprentissage automatique.

PLAN DU MÉMOIRE

Notre travail s'organise autour de quatre (04) chapitres principaux, nous précisons ici l'objectif de chacun d'entre eux avec un bref descriptif de leur contenu.

- Chapitre 1 : Ce chapitre introductif présente une brève introduction au domaine de la cyber-sécurité.
- Chapitre 2 : Dans ce chapitre nous avons présenté les principales techniques de l'apprentissage automatique et son apport dans le domaine de la cyber-sécurité.
- Chapitre 3 : Nous avons présenté dans ce chapitre l'architecture générale et les différents modules de notre solution, nous avons mis aussi en avant les phases nécessaires à la réalisation de notre application.
- Chapitre 4 : Ce chapitre présente l'environnement de développement (matériel et logiciel), les langages de programmation, les librairies et les algorithmes d'apprentissage utilisés ainsi que les résultats obtenus et les tests.
- Conclusion : Enfin, une conclusion vient clôturer ce mémoire.
- Annexes : Cette partie aborde des descriptions détaillées sur quelques notions rencontrées dans les différents chapitres du mémoire.

Chapitre 1

Généralités sur la sécurité informatique

1.1 Introduction

A notre ère, les systèmes informatiques sont omniprésents et interviennent dans pratiquement tous les secteurs d'activités. Ces systèmes sont constitués de différents nœuds (serveurs, ordinateurs, routeurs, etc.) organisés en réseau au sein d'une infrastructure. La complexité croissante de ces systèmes et leur ouverture sur Internet les rendent souvent cible à des attaques malveillantes. Ces attaques peuvent engendrer de sérieux dégâts tels que la perte financière, la nuisance à l'image de l'entreprise, la perturbation de ses opérations critiques, ... etc.
La sécurité des systèmes informatiques est donc devenue un enjeu incontournable et passe par la mise en place de plusieurs moyens et outils afin de prévenir les attaques et réagir contre elles lorsqu'elles sont détectées.
Ce chapitre présente une introduction au domaine de la sécurité informatique. Nous commençons par définir les principaux concepts de la cyber-sécurité, ensuite, nous présentons une classification des attaques selon différents aspects, et enfin, nous abordons les différents moyens et techniques de protection contre les différentes attaques ainsi que leurs limites.

1.2 Concepts de base

La sécurité informatique utilise un vocabulaire bien défini, dans ce qui suit, nous allons présenter les principaux termes liés à la sécurité informatique utilisés dans ce mémoire.

- **Sécurité informatique :** c'est l'ensemble des techniques et outils permettant de garantir trois objectifs essentiels : confidentialité, intégrité et disponibilité. Ces outils peuvent être, organisationnels, matériels, logiciels ou juridiques dont le but est de protéger les informations et les systèmes d'information contre l'accès, l'utilisation malveillante ou non autorisée, la modification, la divulgation et la destruction des données et connaissances [1].

- **Menace :** une menace correspond à une action ou un événement potentiel pouvant compromettre la sécurité des systèmes, réseaux ou équipements numériques [2]. Elle peut aussi bien être intérieure qu'extérieure.
- **Vulnérabilité :** une vulnérabilité est une faille de la sécurité d'un système provoquée au niveau de la conception, de la mise en œuvre ou dans les contrôles internes de ce système. Cela peut être exploité par une personne potentiellement malveillante pour altérer le fonctionnement normal du système ou avoir un accès non autorisé à ses données [3].
- **Attaque :** une action malveillante destinée à porter atteinte à un système informatique. Elle représente la concrétisation d'une menace, et nécessite l'exploitation d'une vulnérabilité [4].
- **Intrusion :** faute malveillante externe résultante d'une attaque qui a réussi à exploiter une vulnérabilité.
- **Risque :** un risque est une évaluation quantifiable du potentiel qu'un événement ou une action cause des dommages ou des pertes aux systèmes d'information [2].
- **Exploit :** un exploit est un type d'attaque qui se présente généralement sous la forme d'un logiciel ou d'un code dont le but est de prendre le contrôle d'un ordinateur.

1.3 Propriétés de la sécurité informatique

Assurer la sécurité d'un système informatique revient à garantir les trois propriétés de la sécurité suivantes : Confidentialité, Intégrité, et disponibilité [5]. Ces dernières sont connues aussi sous le nom de la triade CIA pour Confidentiality, Integrity, Availability en anglais.

- **Confidentialité :** la confidentialité vise à assurer que seuls les sujets (les personnes, machines ou logiciels) autorisés aient accès aux ressources et aux informations auxquelles ils ont droit.
- **Intégrité :** permet d'assurer que les données sont bien celles que l'on croit être. Vérifier l'intégrité des données consiste à déterminer si les données n'ont pas été modifiées ou détruites de façon non autorisée.
- **Disponibilité :** garantir que les informations sont accessibles aux personnes autorisées chaque fois quand elles en ont besoin même pendant les situations d'incidents. Cela implique des mesures telles que la mise en place de redondance et de tolérance aux pannes, la création de plans de sauvegarde et de récupération, et la mise en œuvre de procédures de récupération après incident.

1.4 Les attaques informatiques

Les attaques informatiques peuvent se produire à différentes phases du cycle de vie d'un système : lors de la conception, lors de l'implémentation, lors du déploiement et de la configuration ou bien lors de l'utilisation du système. Dans cette section, nous allons présenter le cycle de vie des attaques, leur classification ainsi que certains exemples d'attaque.

1.4.1 Cycle de vie d'une attaque informatique

En général, une attaque est constituée de plusieurs actions comme la collecte d'informations, l'utilisation de ces informations pour accéder au sein du système dans le but de perturber son fonctionnement. Selon Burgermeister et al [6], le processus d'une attaque informatique est constitué de cinq principales phases :

- **Analyser (Probe) :** consiste à la collecte d'informations sur le système cible par le biais d'un ensemble d'outils, comme par exemple, procéder à un scan de ports en utilisant un programme comme Nmap[1], ou encore, un scan de vulnérabilités à l'aide du programme Nessus[2].
- **Pénétrer (Penetrate) :** il s'agit de l'utilisation des informations récoltées pour s'infiltrer dans le système. Des techniques comme les attaques par dictionnaire peuvent être utilisées pour contourner les protections par mot de passe.
- **Persister (Persist) :** une fois le système infiltré, l'attaquant cherchera à y revenir facilement. L'attaquant peut, par exemple, créer un utilisateur de niveau administrateur avec un mot de passe que lui seul connaît ou installer un logiciel d'accès à distance. La plupart des attaquants tentent de masquer les preuves de leur activité à ce stade en modifiant ou en supprimant les journaux du système et du pare-feu.
- **Propager (Propagate) :** le système est infiltré, l'accès est facile. Le pirate pourra alors explorer le système et trouver de nouvelles cibles qui l'intéresseraient.
- **Paralyser (Paralyze) :** cette étape peut consister en plusieurs actions. L'attaquant peut utiliser le serveur pour mener une attaque sur une autre machine, détruire des données ou encore endommager le système d'exploitation dans le but de causer un dysfonctionnement du serveur.

1.4.2 Classification des attaques informatiques

Les attaques informatiques peuvent être classées de différentes manières selon divers critères, nous mentionnons : la classification selon l'effet, la source, ou la cible de l'attaque.

1. `http://nmap.org/`
2. `http://www.tenable.com/products/nessus`

Classification selon l'effet de l'attaque

- **Les attaques passives :** consistent à accéder, utiliser ou à observer le système cible sans modifier les données ou dysfonctionner les ressources de ce dernier. Elles sont généralement indétectables (par exemple : capture de contenu, analyse de trafic, etc.).
- **Les attaques actives :** consistent à effectuer des changements non autorisés sur les données du système, à s'introduire dans des équipements réseau ou à perturber leur fonctionnement. Les attaques de ce type sont bien évidemment plus dangereuses.

Classification selon la source de l'attaque

- **Les attaques internes :** provenant des employés de l'entreprise ou de leurs partenaires commerciaux ou clients.
- **Les attaques externes :** venant de l'extérieur, fréquemment via Internet.

Classification selon la cible de l'attaque

- **Les attaques réseaux :** les attaques réseaux s'appuient sur des vulnérabilités liées directement aux protocoles réseaux ou à leur implémentation.
- **Les attaques systèmes :** les attaques systèmes s'appuient principalement sur des vulnérabilités spécifiques aux applications utilisées.

1.5 Exemples d'attaques

Dans ce travail, nous allons nous intéresser particulièrement aux attaques ciblant le réseau. Ces dernières utilisent des méthodes qui leur permettent de surveiller le trafic réseau dans le but d'acquérir des informations critiques. Il existe un grand nombre d'attaques réseaux. Nous présentons quelques-unes ci-dessous.

- **Déni de Service (DoS) :** il s'agit d'une attaque visant à saturer les serveurs en utilisant diverses techniques pour les pousser à atteindre leurs limites de ressources (de traitement ou de mémoire) afin de les faire planter. Le but de l'attaque DoS est de rendre le service inaccessible pour les utilisateurs légitimes.
- **Déni de Service Distribué (DDoS) :** Tandis que l'attaque DoS est effectuée par une seule machine, l'attaque DDoS est une variante de l'attaque DoS dont la principale différence est que l'attaque est effectuée par un réseau de systèmes informatiques infectés et contrôlés par l'attaqueur, aussi appelé : Réseau zombie ou botnet[3]. Parmi les programmes utilisés pour effectuer des attaques DoS et DDoS sont : Slowloris, HULK et GoldenEye.

3. Le botnet (réseau zombies) constitue l'ensemble des machines compromises sur lesquelles sont installées des composants malveillants et autonomes.

- **Les attaques web :** c'est des attaques qui exploitent des vulnérabilités présentent dans les applications web, voici quelques exemples d'attaques web :
  - **L'attaque par force brute :** c'est une attaque par approche essai-erreur en testant les combinaisons possible afin de deviner un mot de passe, un nom d'utilisateur, ou afin de cracker une clé de chiffrement, etc.
  - **L'injection SQL :** c'est une attaque qui exploite les vulnérabilités sur les champs d'entrée d'une application web qui transmettent des informations directement vers la base de données, et cela a pour but d'injecter du code SQL malicieux pour gagner un accès non autorisé à la base de données. L'injection SQL est une attaque qui opère du côté serveur.
  - **Le Cross-Site Scripting (XSS) :** c'est un script malicieux écrit en langage(s) client injecté dans les données fournies par l'utilisateur. Cette injection se fait généralement à travers les formulaires, les hyperliens dans le but d'être exécutée sur le navigateur de la machine victime. L'attaque XSS opère du côté client.

1.6 Monitoring de la cybersécurité

Le succès de toute stratégie de cybersécurité est proportionnel à la quantité d'informations collectées sur les potentielles menaces et l'efficacité de leur analyse afin de les détecter au moment opportun, d'où l'importance de la surveillance permanente. Dans ce qui suit, nous allons introduire les systèmes de surveillance (monitoring) de la cybersécurité et les outils utilisés pour les mettre en œuvre.

1.6.1 Définition du monitoring du cybersécurité

Un système de monitoring de la cybersécurité est un outil qui permet de surveiller en temps réel les activités sur les réseaux et les systèmes informatiques, afin de détecter les comportements anormaux ou malveillants, les attaques potentielles, les vulnérabilités de sécurité et les tentatives de piratage. Les données collectées sont stockées dans des fichiers journaux ou des bases de données pour une analyse ultérieure et peuvent déclencher des alertes pour signaler des activités suspectes. L'objectif principal est d'aider les organisations à comprendre et à atténuer les risques liés aux menaces inconnues (Zero-Day Attacks[4], les vulnérabilités Forever-day[5]) et organisés (APT[6]) [7].

4. Les attaques zero-day sont des attaques qui exploitent les vulnérabilités de sécurité encore inconnues et pour lesquelles il n'existe pas de correctif ou de solution de contournement connue.
5. Les vulnérabilités "Forever-day" sont similaires aux attaques "zero-day", mais plutôt que d'être des failles de sécurité inconnues, elles sont connues des attaquants depuis longtemps et ne sont pas corrigées, offrant ainsi un accès continu et illimité à un système vulnérable.
6. Les APT (Advanced Persistent Threat) sont des menaces sophistiquées qui cherchent à accéder à des données sensibles en restant indétectables. Les attaquants sont hautement qualifiés et peuvent rester infiltrés pendant longtemps.

1.6.2 Processus du monitoring de la cybersécurité

Le processus des systèmes de monitoring pour la cyber-sécurité peut être décomposé en plusieurs étapes (cf. figure 1.1) [8] :

- **Collecte de données :** les systèmes de monitoring collectent des données sur les activités depuis les réseaux informatiques et les systèmes, tels que les adresses IP, les ports réseau, les noms d'utilisateurs, les noms de fichiers et les types de fichiers, les activités de navigation web, les comportements d'attaques malveillantes, les tentatives d'intrusion, les échanges de données et les communications entre systèmes.
- **Stockage de données :** les données collectées sont stockées dans des fichiers journaux ou des bases de données, qui peuvent être analysées ultérieurement pour détecter les comportements anormaux ou malveillants.
- **Analyse de données :** les données stockées sont analysées pour détecter les activités suspectes ou malveillantes. Cela peut impliquer l'utilisation d'algorithmes et de techniques d'analyse des données pour identifier des modèles dans les données et les comparer avec des modèles de comportement typiques ou connus pour détecter les anomalies.
- **Génération d'alertes :** lorsque des comportements anormaux ou malveillants sont détectés, des alertes sont déclenchées pour signaler ces activités aux administrateurs de sécurité. Ces alertes peuvent être classées en fonction de leur gravité ou de leur priorité, ce qui permet aux équipes de prioriser leurs actions de réponse. Elles peuvent être envoyées par courrier électronique, SMS ou notifications sur des tableaux de bord.
- **Rapports et analyses :** les systèmes de monitoring pour la cybersécurité peuvent également générer des rapports et des analyses pour permettre aux administrateurs de sécurité de comprendre les tendances de sécurité, les vulnérabilités de leurs systèmes et les performances de leurs outils de sécurité.

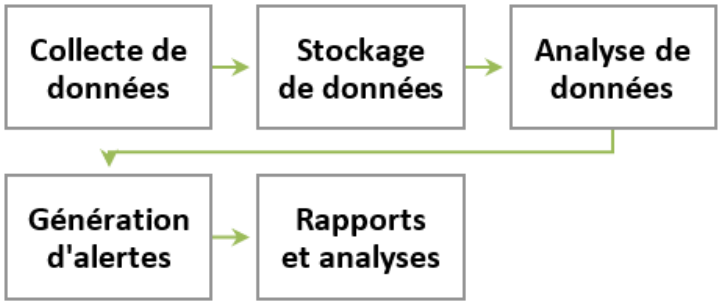

FIGURE 1.1 – *Processus de monitoring du Cyberespace*

Il est important de noter que le processus du monitoring de la cybersécurité est itératif et continu. Il nécessite une vigilance constante et une adaptation aux nouvelles menaces et techniques d'attaques.