Intégration de la CTI et de l'apprentissage automatique pour une détection améliorée des menaces numériques

Projet M2 pour l'obtention du « titre d'ingénieur diplômé de l'Institut Supérieur d'Electronique et du Numérique – Junia »

Projet 03

ISEN M2

2023-2024

Tanguy SINGEOT-SOUSA	M2 – Intelligence artificielle
Loïc BLONDEAU	M2 – Cybersécurité
Cléo DEMAY	
Arthur FAGOT	
Théo WATTEL	

Enseignante encadrante	Mounia ZAYDI
Date de la présentation	31/05/2024

Résumé

Les récents progrès en matière d'apprentissage automatique et d'Intelligence Artificielle (IA) poussent la recherche actuelle à se focaliser sur son utilisation dans le cadre de la détection de malware. Une fusion judicieuse et l'intégration de données de la Cyber Threat Intelligence (CTI) pourraient permettre une meilleure standardisation des données, une collecte simplifiée et une exploitation plus rapide pour une détection améliorée. La performance de ces solutions dépend également de la capacité de ces techniques à suivre l'évolution des logiciels malveillants et à s'adapter aux nouvelles technologies. L'objectif de ce mémoire de projet est de faire un état de l'art de la recherche actuelle sur les thèmes de l'IA, la détection de malware et les données de la CTI afin d'en faire une analyse critique et de concevoir une solution innovante pour renforcer la cybersécurité en exploitant les avancées actuelles.

Mots-clés : Logiciel malveillant, Cybersécurité, Intelligence Artificielle, Apprentissage automatique, Cyber Threat Intelligence.

Abstract

Recent advances in machine learning and artificial intelligence (AI) are pushing current research to focus on its use in malware detection. Judicious fusion and integration of Cyber Threat Intelligence (CTI) data could enable better data standardization, simplified collection, and faster exploitation for improved detection. The performance of these solutions also depends on the ability of these techniques to keep up with the evolution of malware and adapt to new technologies. The objective of this document is to make a state of the art of current research on the topics of AI, malware detection and CTI data in order to make a critical analysis and design an innovative solution to strengthen cybersecurity by exploiting current advances.

Keywords : Malware, Cybersecurity, Artificial Intelligence, Machine Learning, Cyber Threat Intelligence.

Remerciements

Nous tenons à remercier chaleureusement la commanditaire et encadrante du projet, Madame Mounia ZAYDI pour sa disponibilité, la pertinence de ses conseils et la richesse des échanges que nous avons pu avoir. Tant sur les aspects techniques, documentaires ou organisationnels, elle nous a suivi avec attention tout au long du déroulement du projet.

Par ailleurs, nous sommes particulièrement reconnaissants envers Madame Pascale DIENER et Monsieur Gabriel CHÊNEVERT, membres de l'équipe pédagogique et administrative de Junia, notamment pour le suivi des projets ainsi que pour leur aide logistique.

Plus généralement, c'est toute l'équipe que nous souhaitons remercier pour leur sérieux et leur investissement.

Glossaire

A

API

Une API, ou Application Programming Interface, désigne l'ensemble de classes, méthodes, fonctions et constantes offert par un logiciel et qui peut être utilisé par un autre logiciel.

C

Clé de registre

Une clé de registre est une petite unité de donnée qui stocke des informations sur les préférences utilisateur, la configuration, les appareils.

Cyberdéfense

La cyberdéfense correspond à l'ensemble des moyens mis en place dans un pays dans le cadre de la guerre informatique menée dans le cyberespace.

Cyberespace

Le cyberespace est l'ensemble des équipements, des applications et des données numériques dans le monde.

Cybersécurité

La cybersécurité est un domaine informatique dont le but est de protéger les systèmes, réseaux et programmes contre les attaques numériques.

D

Dataset

Un jeu de donnée, ou dataset, est un regroupement de données liée de manière cohérente.

E

Endpoint

D'après Microsoft, les endpoints dits points de terminaison en français sont des appareils physiques qui se connectent à un réseau informatique et échangent des informations avec lui.

ESN

(Entreprise de Services du Numérique) Entreprise experte dans le domaine du numérique ayant pour objectif d'accompagner des clients dans la réalisation d'un projet. Elle peut englober plusieurs métiers tel que le conseil, la conception ou encore la formation.

H

Honeypot

Un honeypot est un système (par exemple, un serveur Web) ou une ressource système (par exemple, un fichier sur un serveur) conçu pour attirer les pirates et les intrus potentiels, comme le miel attire les ours.

Hyperviseur

Un hyperviseur est un logiciel qui crée et gère des machines virtuelles.

I

Indicateurs clé de performance

Un indicateur clé de performance, ou key performance indicator en anglais, est un indicateur utilisé pour l'aide à la décision. Dans le cadre du machine learning, il permet d'évaluer les performances d'un ou plusieurs modèles.

Intelligence Artificielle

L'intelligence artificielle (IA) regroupe l'ensemble des techniques dont le but est de créer des machines capables de simuler l'intelligence humaine.

Intrusion Detection System

Un IDS est un dispositif ou logiciel de sécurité qui surveille les réseaux ou systèmes pour détecter des activités malveillantes ou des violations de politiques de sécurité.

L

Large Language Model

Sous domaine de l'IA, modèle capable de prédire quel mot est le plus probable après un autre, construisant une suite de mots séquentiellement sur cette base.

M

Machine virtuelle

Machine avec un système d'exploitation différent partageant le matériel physique de la machine hôte tout en restant isolée de celle-ci.

Malware

Un malware, ou logiciel malveillant, est un programme hostile déstiné à éxecuter des instructions sur un système informatique sans le consentement de l'utilisateur.

Mutex

Un mutex est une primitive de synchronisation utilisée en programmation informatique pour éviter que des ressources partagées d'un système ne soient utilisées en même temps.

N

Notebook

Les notebooks Jupyter sont des cahiers électroniques qui, dans le même document, peuvent rassembler du texte, des images, des formules mathématiques et du code.

P

Payload

Le payload, ou la charge utile en français, est la partie d'un élément transmis qui contient des données à différencier des autres données, comme les données d'acheminement. Dans le cadre d'une attaque, le payload malveillant est ce qui contient le maliciel, il peut être exécuté automatiquement ou sur action de l'utilisateur.

S

Sandbox

Une sandbox est un environnement isolé utilisé pour exécuter et analyser des programmes potentiellement malveillants sans risque pour le système hôte.

Secure coding

Le secure coding dit aussi programmation sécurisée consiste à prendre en compte la sécurité informatique à tous les moments de la conception, de la réalisation et de l'utilisation d'un programme informatique. [16]

Security Operation Center

Plateforme permettant la supervision et l'administration de la sécurité du système d'information au travers d'outils de collecte, de corrélation d'événements et d'intervention à distance.

Shell code

Un shellcode est une chaîne de caractères qui représente un code binaire exécutable. À l'origine destiné à lancer un shell, le mot a évolué pour désigner tout code malveillant qui détourne un programme de son exécution normale. [15]

Abréviations

- **ANN :** Artificial Neural Network
- **CERT :** Computer Emergency Response Team
- **CKC :** Cyber Kill Chain
- **CRISP-DM :** Cross-Industry Standard Process for Data Mining
- **CTI :** Cyber Threat Intelligence
- **DDoS :** Distributed Denial-of-Service
- **DL :** Deep Learning
- **DLL :** Dynamic Link Library
- **EDR :** Endpoint Detection and Response
- **ESN :** Entreprise de Services Numériques
- **IA :** Intelligence Artificielle
- **ICOSEC :** International Conference on Smart Electronics and Communication
- **IDS :** Intrusion Detection System
- **IOC :** Indicator Of Compromise
- **KNN :** K-Nearest Neighbors
- **KVM :** Kernel-based Virtual Machine
- **MDR :** Managed Detection and Response
- **ML :** Machine Learning
- **OASIS :** Organization for the Advancement of Structured Information Standards
- **OS :** Operating System
- **OSINT :** Open Source INTelligence
- **SDOs :** STIX Domain Objects
- **SDROs :** STIX Relationships Objects
- **SOC :** Security Operation Center
- **STIX :** Structured Threat Information eXpression
- **TAXII :** Trusted Automated Exchange of Intelligence Information
- **TIC :** Technologies de l'Information et de la Communication
- **TTP :** Tactiques, Techniques et Procédures
- **UKC :** Unified Kill Chain
- **VM :** Virtual Machine
- **XDR :** eXtended Detection and Response
- **XGBoost :** eXtreme Gradient Boosting

Sommaire

1. Introduction

De nos jours, la cybersécurité représente un défi crucial pour toutes les entreprises, qui font face à des menaces de plus en plus complexes et fréquentes. La multiplication des attaques cybernétiques souligne l'importance capitale de mettre en place des défenses solides et proactives afin de préserver les systèmes d'information.

Popularisé en novembre 2022, le Large Language Model (LLM) ChatGPT développé par OpenAI a permis au grand public de réaliser la puissance de l'Intelligence Artificielle.

L'IA est un processus qui vise à imiter l'intelligence humaine au travers d'algorithmes informatiques [1]. Elle donne à la machine la capacité d'analyser et de prendre des décisions. Au sein de l'IA, on distingue généralement deux sous-disciplines basées sur l'apprentissage que sont le Machine Learning (ML) et le Deep Learning (DL).

Bien que l'IA puisse transformer le travail des usagers pour améliorer leur productivité avec pour exemple Copilot pour Microsoft 365 [2], du fait de ses multiples capacités, elle peut également être utilisée à mauvais escient. L'IA est à la fois capable de générer du code malveillant ou d'être utilisée à des fins d'ingénierie sociale en imitant les comportements humains dans le cadre de la rédaction de mail de phishing par exemple. Au-delà d'une utilisation détournée des capacités de l'IA, les menaces cybernétiques se développent et se complexifient à grande vitesse. En 2023, on répertorie 28 902 Common Vulnerabilities and Exposures (CVEs) publiées, ce qui représente une hausse de 15% par rapport à l'année 2022 [3].

Face à l'augmentation des cyberattaques, la sécurisation des systèmes d'information est devenue une priorité. La Cyber Threat Intelligence, est une discipline qui gagne en importance et devient essentielle. La CTI est une discipline basée sur l'analyse des menaces cybernétiques. Elle fournit aux analystes des renseignements contextualisés permettant de comprendre le mode opératoire des attaquants. Des données de CTI partagées permettent une réaction proactive aux menaces et une prise de décision rapide lors d'une crise.

Puisque la cyberdéfense doit évoluer au moins aussi rapidement que les cybermenaces, l'utilisation de données de la CTI et l'appropriation de l'IA à des fins de défense et de détection sont en plein essor. Les liens entre IA et cybersécurité se complexifient, comme peut en

témoigner le thème du Forum InCyber[1] 2024 : « Ready for AI ? Réinventer la cybersécurité à l'air de l'IA » [4]. Selon la Commission Nationale de l'Informatique et des libertés (CNIL)[2], la cybersécurité « représente les technologies pour protéger son patrimoine informationnel et les personnes concernées des atteintes à leur données » [5].

Combiner IA et cybersécurité est le sujet de nombreuses recherches depuis plusieurs années. Certaines solutions apparaissent sur le marché mais le sujet n'en est qu'à ses débuts. Les services proposés utilisent souvent l'IA, et plus précisément le ML, pour détecter les comportements anormaux. Cependant les liens avec la CTI restent sommaires. Ce projet vise alors à élaborer une question de recherche claire et à proposer une solution innovante pour renforcer la cybersécurité en exploitant les avancées actuelles en CTI et en IA.

Ce projet s'inscrit donc dans le cadre de l'obtention du « titre d'ingénieur diplômé de l'Institut Supérieur d'Electronique et du Numérique – Junia » et constitue un module important du master 2 de la formation Junia ISEN.

Les objectifs sont multiples. Afin d'élaborer une question de recherche précise, il est nécessaire de synthétiser un grand nombre de recherche sur les sujets inhérents à ce projet au travers d'une revue de littérature. Cela constitue une première approche de la recherche scientifique et une introduction à un sujet de thèse. Une autre mission consiste à renforcer la posture de sécurité des organisations face aux cybermenaces en tirant partie de la puissance de la CTI et de l'apprentissage automatique. A ce titre, la comparaison de différents modèles et la conception d'un prototype de détection représentent des objectifs supplémentaires.

Ce mémoire de projet s'organise en quatre parties. Dans un premier temps, le cadre du projet ainsi que les méthodologies employées sont définis. Ensuite, l'état de l'art des connaissances sur les différents sujets permet une meilleure approche du projet. La partie recherche qui s'en suit développe la composition de la revue de littérature, présente une analyse critique de l'existant et détaille l'élaboration de la question de recherche ainsi que la solution envisagée. Enfin, la partie réalisation qui précède une conclusion générale, expose le travail technique effectué, les outils et méthodologies utilisés et les résultats obtenus.

[1] Le Forum InCyber (FIC) est un évènement européen annuel sur les questions de la sécurité et de la confiance numérique. [60]

[2] La CNIL est « le régulateur des données personnelles, elle accompagne les professionnels dans leur mise en conformité et aide les particuliers à maitriser leurs données personnelles et à exercer leurs droits » [61].

2. Cadrage du projet

2.1. Composition de l'équipe

L'équipe en charge de ce projet est composée de cinq étudiants en master 2 Junia ISEN. Quatre d'entre eux sont spécialisés en cybersécurité : Loïc BLONDEAU, Cléo DEMAY, Arthur FAGOT et Théo WATTEL. Spécialisé en Intelligence Artificielle, Tanguy SINGEOT-SOUSA, vient compléter l'équipe. La variété des profils a permis des échanges riches et un travail de qualité. L'ensemble du projet a été encadré par Madame Mounia ZAYDI, membre du corps enseignant de Junia ISEN (voir *Figure 1*).

Figure 1- Organigramme des membres de l'équipe

2.2. Choix de la thématique

Le choix d'un projet orienté vers la sécurité informatique côté défensive représente pour la majorité des membres du groupe un lien direct avec le métier exercé en entreprise lors du contrat de professionnalisation. En effet, une compréhension du terrain avec par exemple la connaissance du fonctionnement d'un Security Operation Center (SOC) a permis d'appréhender rapidement les différents concepts et de visualiser les potentielles applications réelles du projet. Se perfectionner sur le domaine assez spécifique de la CTI représente une réelle opportunité de spécialisation future.

De la même manière, la présence d'un membre spécialisé en Intelligence Artificielle a pour intérêt de décupler les compétences techniques au sein du groupe. La mise en commun des différentes connaissances est extrêmement bénéfique.

Parmi les différents objectifs du sujet initial, le présent projet cible la thématique de la détection des malwares qui représentent aujourd'hui un défi important de la cyberdéfense[3]. En effet, entre 2022 et 2023, une augmentation de 23,8% de signalement de ransomware a été observée par l'ANSSI [6]. D'après le National Institute of Standards and Technology, un

[3] Cyberdéfense : « Ensemble des moyens mis en place par un État pour défendre dans le cyberespace les systèmes d'information jugés d'importance vitale, qui contribuent à assurer la cybersécurité » [52].

malware peut être défini comme tel : « Programme implanté dans un système, généralement de manière clandestine, dans le but de compromettre la confidentialité, l'intégrité ou la disponibilité des données, des applications ou du système d'exploitation de la victime, ou de la gêner ou de la perturber d'une manière ou d'une autre. » [7]. En s'implantant dans un système, un malware laisse des marquants sur son passage. Ces derniers sont des indicateurs de compromissions permettant aux analystes de détecter la présence d'anomalies et de qualifier le comportement exact du logiciel malveillant.

La Cyber Threat Intelligence est une discipline basée sur l'analyse des menaces cybernétiques. Elle regroupe l'étude de ces marquants, des méthodes d'attaque utilisées par les pirates informatiques et de l'exploitation des vulnérabilités présentes dans les systèmes d'information actuels.

Du fait de ses évolutions récentes, les liens entre l'Intelligence Artificielle et la cybersécurité deviennent de plus en plus étroits et interdépendants. L'IA permet alors l'exploitation et l'enrichissement des données de la CTI dans le cadre de la détection de malware, ce qui en fait un sujet d'étude pertinent pour ce projet.

2.3. Objectifs du projet

Les attentes pour ce projet sont multiples et ont évoluées au cours des semaines de projet. Bien que le sujet initial soit un travail de recherche, des résultats techniques ont ensuite été attendus.

Puisque ce projet représente une introduction à l'exercice de la thèse, la majorité du travail attendu consiste à définir une problématique claire qui se décline en question de recherche pertinente. A ce titre, la rédaction d'une revue de littérature composée de fiches de lecture centrées sur une recherche bibliographique détaillée, représente un objectif essentiel. Un travail autonome et rigoureux du groupe, une capacité d'analyse, d'interprétation et de communication des données collectées – à l'oral ou à l'écrit – de façon claire et concise représentent tout autant d'objectifs secondaires inhérents à ce type de travail. La formulation d'une ébauche de réponse à notre question de recherche constitue un objectif supplémentaire.

Concernant la partie réalisation, l'objectif est de s'appuyer sur la partie recherche documentaire pour lister un certain nombre de modèles judicieux pour notre étude afin de les comparer et de mesurer leur pertinence. L'appropriation et l'utilisation d'une sandbox afin de faire de la détonation et de la récupération de logs représente l'un des défis de ce projet.

L'extraction des données de CTI permettront l'obtention de données exploitables par les modèles de détection étudiés.

Étant donné l'ampleur du sujet, il a été précisé qu'un livrable technique complètement fonctionnel, intégrable dans un système d'information n'était pas l'attente minimale. En effet, l'enjeu principal est la définition d'un modèle d'Intelligence Artificielle innovant permettant d'améliorer significativement la détection de malware permettant ainsi aux organisations de renforcer leur cybersécurité et de réduire leur exposition aux risques.

2.4. Méthodologie

L'approche agile a été utilisée dans le cadre de ce projet afin d'atteindre les objectifs. L'application exacte d'une méthodologie connue a été compliquée compte tenu de la partie recherche du projet et du découpage du temps de travail au fil de l'année. Une approche inspirée des méthodes préexistantes Scrum [8] et Kanban [9] a été mise en place.

L'approche combinée utilisée dans ce projet est composée de différents sprints de travail de durée inégales contrairement à ce que préconise la méthode Scrum, ces derniers seront détaillés dans la partie suivante. Chaque sprint commence par une réunion de démarrage qui définit les objectifs de la période, les tâches allouées à chacun et planifie les points réguliers de l'équipe avec ou sans la commanditaire. Le backlog de tâches pour la période est ainsi mis à jour avec des priorités et des dates limites. Lors des périodes de travail, chacun a le loisir de travailler en autonomie et doit rendre régulièrement compte de son avancement. A la fin du sprint, une réunion de « review » est organisée pour faire le bilan de la période, l'état de l'avancement et définir ce qui aurait pu être mieux effectué. A la suite de cette réunion, le compte rendu de la période est rédigé et envoyé au commanditaire.

La méthode Kanban a également inspiré l'approche utilisée pour ce projet au travers de l'outil Notion[4] [10]. Ce dernier a été utilisé afin d'obtenir des représentations visuelles claires de l'avancement du projet. Ainsi, la gestion des flux a été améliorée grâce au tableau d'avancement qui centralise les informations mises à jour continuellement en fonction du développement du projet.

L'avantage de cette approche est sa flexibilité qui a permis à l'équipe de s'adapter à tous les aléas ou changements d'attentes tel que l'intégration d'une partie réalisation au travail de

[4] Notion est une application de prise de notes, de gestion de projet et de collaboration développée par Notion Labs Inc.

recherche initial. À la suite de cette modification, il a été nécessaire d'adapter la gestion de projet. La partie technique s'est avérée plus complexe et chronophage que prévu. La méthode CRISP-DM (Cross-Industry Standard Process for Data Mining) [11] a permis un avancement concret et l'obtention de résultats tangibles. Son utilisation sera détaillée dans la partie réalisation de ce rapport.

Enfin, la répartition judicieuse des tâches en fonction des compétences et appétences de chacun a permis une avancée efficace sur différents sujets en parallèle. Un soin particulier a été porté à la communication au sein des membres afin de garantir l'homogénéité du projet.

2.5. Planification des activités

Afin d'atteindre les objectifs fixés, une gestion de projet rigoureuse et méthodique a été mise en place. A la suite de la présentation des projets, un grand brainstorming a été organisé afin de regrouper les idées de chacun et commencer à organiser la gestion du temps.

Le temps et les horaires de travail ont été fixés en fonction des emplois du temps de chacun. Lors des semaines entières consacrées au projet, la durée de travail d'une journée a été fixée à 8h afin de permettre de la réalisation du nombre d'heure exigées (332h/Homme). Généralement, les horaires sont 8h30-12h puis 13h30-18h, elles représentent 1h de plus que demandées par l'administration de l'école mais sont nécessaires à la bonne réalisation du projet. En dehors des semaines dédiées et des heures planifiées par l'école, chacun est libre de planifier ses heures comme bon lui semble mais un tableur Excel regroupant les temps de travail et activités de chacun a été mis en place afin garder un œil sur le respect des horaires. En pratique, la durée de projet a été de 11 semaines de période école et de 4 semaines dédiées au projet.

Afin de faciliter la communication et le partage de documents, une équipe Teams a été créée. Un point de suivi avec la commanditaire a également été organisé environ toutes les deux semaines en période école. En plus de ces points de suivis, lors des longues périodes école, des réunions hebdomadaires ont été mises en place et lors des semaines complètes de projet, l'équipe s'est réunie deux fois par semaine.

L'outil privilégié pour la gestion de ce projet est Notion et son utilisation et son organisation visuelle a évolué durant l'année. Il regroupe un journal de bord comprenant la définition du projet et une page mensuelle dans laquelle des notes et comptes-rendus de réunions sont consignés. L'onglet « Gestion de projet » a été le plus utile car il a permis le pilotage du projet tout au long de l'année. On y retrouve notre diagramme de Gantt mis à jour de manière

hebdomadaire ainsi qu'un tableau Kanban de gestion des tâches. Il est ainsi possible d'ajouter différentes tâches, de les assigner, définir une date d'échéance, une priorité et un statut (to do, in progress, achevée, en retard). Un nouvel onglet « Gestion de projet mai » a été ajouté afin de piloter au mieux les dernières semaines du projet.

Toujours dans un souci d'efficacité et d'optimisation du temps, des templates ont été élaborés pour les documents importants du projet (comptes-rendus mensuels et hebdomadaires, fiches de lecture ou notes d'informations).

Le découpage temporel du projet a été définis dès fin septembre et a été adapté en fonction des avancés de l'équipe (voir *Tableau 1*).

Tableau 1 - Découpage temporel du projet

Période de projet	Découpage temporel
1ère période école : 19/09/2023 – 5/11/2024	• Découverte du projet • Recherches préliminaires
1ère semaine de projet : 8/01/2024 – 14/01/2024	• Premières fiches de lecture • Critique des fiches de lectures • Elaboration d'une première question de recherche.
2ème période école : 15/01/2024 – 25/02/2024	• Division de l'équipe • Début de la partie technique • Enrichissement de la revue de littérature • Précision de la question de recherche • Point de mi-parcours.
2ème semaine de projet : 13/05/2024 – 19/05/2024	• Rédaction de la première partie du rapport • Finalisation de la collecte de logs • Tests d'algorithmes avec de nouveaux jeux de données

3ème semaine de projet : 20/05/2024 – 26/05/2024	• **Analyse et optimisation de différents algorithmes d'apprentissage** • **Rédaction de la partie recherche du rapport** • **Rédaction de la partie technique du rapport** • **Ecriture de la vidéo d'introduction**
4ème semaine de projet : 27/05/2024 – 31/05/2024	• Finalisation du rapport • Tournage et montage de la vidéo d'introduction • Préparation de la soutenance • Pré-soutenance • Soutenance finale

3. Etat de l'art : concepts fondamentaux, recherche et apprentissage

3.1. La Cyber Kill Chain : modéliser les cyberattaques[5]

3.1.1. Définition de la Cyber Kill Chain (CKC) ?

La "Cyber Kill Chain" (chaîne de cyber-attaques) est un concept développé par la société de sécurité informatique Lockheed Martin [12] pour décrire les étapes successives qu'une attaque informatique typique suit, du stade initial de la planification jusqu'à la compromission finale de la cible. Il existe différents modèles de Cyber Kill Chain. Cette approche a été élaborée pour aider les organisations à comprendre et à se défendre contre les cyber-attaques en décomposant le processus en phases distinctes. Le processus identifie toutes les étapes que l'attaquant doit effectuer pour atteindre son objectif.

3.1.2. La Cyber Kill Chain Lockheed Martin

Le modèle Lockheed Martin [13] est composé de 7 étapes. L'attaquant réussit son intrusion si et seulement s'il parvient à réaliser les 6 premières étapes de la Kill Chain.

[5] Toutes les informations présentes dans cette partie proviennent de la note d'information « Note d'informations – Cyber Kill Chain » rédigée le 10 octobre 2023 par Cléo DEMAY dans le cadre de la rédaction de la revue de littérature inhérente à ce projet (voir *Annexe 25*).

3.1.2.1. Les étapes de la Kill Chain

Repérage (RECONNAISSANCE) : Identifier la cible

- L'attaquant se renseigne sur l'organisation de l'entreprise, recherches sur les réseaux sociaux des employés, découverte des serveurs disponibles sur internet…
- Cette étape peut être difficile à repérer mais sa détection, même après coup, permet de diagnostiquer les intentions de l'attaquant.
- Une solution peut être la mise en place de règles de détection des comportements propres à la reconnaissance (scan de port, analyse des journaux de serveurs web pour détecter des requêtes suspectes, analyse des entêtes http, surveillance des tentatives d'authentification, analyse des robots d'exploration).

Militarisation (WEAPONIZATION) : Préparer l'opération

- L'attaquant créé un vecteur d'attaque afin d'exploiter une vulnérabilité connue ou méconnue, découverte lors de la phase de repérage.
- L'arme est souvent composée d'un malware et d'un exploit dans un payload qui pourra être livré.
- La compréhension de cette phase est essentielle pour les équipes de cyberdéfense. Même s'il est difficile de la détecter au moment où elle se produit, il est possible d'analyser les artefacts de malware et d'en déduire la manière dont ils ont été construits/exploités.

Livraison (DELIVERY) : Lancement de l'opération

- Le vecteur d'attaque est envoyé à la cible via différent moyens : phishing, USB, sites web compromis ou encore interaction sur les réseaux sociaux.
- Côté cyberdéfense, c'est la première et meilleure opportunité pour bloquer l'attaque.

Exploitation (EXPLOITATION) : Accéder à la victime

- Le logiciel malveillant est exécuté avec succès sur la cible, exploitant une vulnérabilité
- Pour éviter cette phase, la formation et la sensibilisation des employés est essentielle.
- Il faut durcir la sécurité des machines (restreindre les accès administrateurs par exemple ou créer des règles personnalisées pour bloquer l'exécution de Shell code)
- Il faut mettre en place des mesures de secure coding.

Installation (INSTALLATION) : Etablir un point d'ancrage

- Généralement, l'attaquant installe une backdoor pour maintenir un accès.
- Côté cyberdéfense, il faut surveiller les installations au moment de l'attaque pour ensuite pouvoir les supprimer.

Command and control (C2) : Contrôle à distance

- Le logiciel malveillant ouvre une console afin de contrôler et manipuler à distance la victime.
- Du côté de la défense, la dernière chance pour bloquer l'attaque est de couper ce canal de communication.
- L'analyse du logiciel permet de découvrir le C2 et ainsi de consolider le réseau pour éviter cette possibilité.

Actions sur les objectifs : Réussir l'objectif

- Si l'attaquant parvient à effectuer toutes les étapes précédentes, il est possible pour lui d'atteindre son but.
- Les objectifs dépendent de l'attaquant en question : récupération d'informations d'identification, escalade de privilège, exfiltration ou modification de données.
- Plus longtemps l'attaquant possède l'accès à l'endpoint, plus grave peut être l'attaque. Il est possible de mettre en place des honeypots afin de détecter au plus vite cette étape et de mettre en place un plan de réponse à incident et de reprise des activités.

3.1.3. La Kill Chain MITRE ATT&CK

MITRE est une organisation à but non lucratif américaine dont l'objectif est de faire progresser la sécurité nationale et servir l'intérêt public [14]. ATT&CK est une base de connaissances accessible à tous d'attaques et de techniques. Cette base de connaissances est utilisée afin de développer des modèles et méthodologies de menaces spécifiques.

Le système ATT&CK comporte 3 matrices principales : Enterprise, Mobile et ICS (Système de Contrôle Industriel). Les Kill Chain diffèrent en fonction de la matrice et de l'OS utilisé ce qui les rendent très complètes. Dans chaque matrice et sous chaque étape, sont regroupées différentes techniques et sous-techniques.

Les différentes phases de la Kill Chain MITRE ATT&CK pour les entreprises sont les suivantes :

1. Reconnaissance ;
8. Credential access ;

2. Resource development ;
3. Initial access ;
4. Execution ;
5. Persistence ;
6. Privilege escalation ;
7. Defense evasion ;
9. Discovery ;
10. Lateral movement ;
11. Collection ;
12. Command and control ;
13. Exfiltration ;
14. Impact.

Sur le site de MITRE, il est également possible de consulter les différentes tactiques et techniques répertoriées. Les tactiques représentent le « pourquoi », ce qui motive l'attaquant à faire son choix alors que les techniques explicitent le « comment ». La catégorie « mitigations » quant à elle représente des concepts de sécurité et des technologies qui peuvent être utilisées pour empêcher l'exécution réussie d'une technique ou d'une sous-technique.

3.1.4. Existence d'autres Cyber Kill Chain

Bien que Lockeed Martin soit précurseur, différentes organisations ont construit leurs propres Kill Chains pour essayer de modéliser au mieux différentes menaces, il en existse alors une multitude. Il est possible de citer pour exemple la Cyber Kill Chain Unifiée (voir *Annexe 18*- Fiche de lecture 8). Ce modèle élaboré par Paul POLS dans son livre blanc « The Unified Kill Chain, Raising resilience against advanced cyber attacks » [15] vise à être plus complet que les modèles traditionnels de Cyber Kill Chain afin de faire face aux attaques les plus avancées. Le document traite de la nécessité de comprendre et de défendre contre les cyberattaques avancées dans un contexte de dépendance croissante à l'égard des Technologies de l'Information et de la Communication (TIC). Le modèle traditionnel de la "Cyber Kill Chain" de Lockheed Martin est mentionné comme une norme de l'industrie pour la modélisation des menaces, mais il est critiqué pour son accent sur les périmètres de sécurité et les attaques de malware. L'article explique alors qu'un modèle plus complet est nécessaire pour faire face aux attaques plus avancées. Ainsi naît l'idée de la "Unified Kill Chain" (UKC), un concept qui vise à modéliser les cyberattaques de bout en bout, depuis les premières étapes de l'attaquant jusqu'à l'atteinte de son objectif.

3.2. Les différents types de maliciels et leurs impacts sur les systèmes

3.2.1. Qu'est-ce qu'un maliciel ?

Dans un contexte géopolitique tendu et marqué par de nombreuses tentatives de déstabilisation d'événements internationaux, le niveau de la menace ne cesse d'augmenter sur le sol français comme à l'international. Selon l'ANSSI, l'espionnage stratégique et industriel représente ce qui a mobilisé le plus de temps aux équipes, l'agence note une augmentation significative du ciblage d'entités travaillant dans des domaines stratégiques, la transmission de données sensibles et la fourniture de services numériques (ESN) [6]. C'est dans ce contexte qu'un grand nombre de maliciels ont vu le jour et les techniques d'infection ont été perfectionnées.

A la base simplement intégré comme un cheval de Troie dans un téléchargement internet ou via un payload intégré dans une clé USB, les méthodes d'insertion de virus dans les systèmes d'information ont considérablement évolué. En effet l'essor des XDR[6], EDR[7], MDR[8] ou plus simplement les antivirus classiques pour les particuliers ont rendu ce type de vecteur d'attaque obsolète et très facilement détectable. Les maliciels (principalement de type virus) inventoriés par les fournisseurs de solutions antivirales voient leur signature stockée dans une base de données. Ces bases de données de signatures sont mises à jour régulièrement mais ne peuvent donc que détecter des menaces connues. Cela ne pose pas de problème spécifique jusqu'au point ou la menace se multiplie tellement que la demande va évoluer vers la détection de maliciels inconnus par les bases de données de signatures.

Face à ces évolutions, les outils d'analyse ont subi des évolutions importantes avec l'arrivée de l'analyse en temps réel des programmes. Deux méthodes voient le jour : d'un côté l'analyse heuristique qui va décompiler le code source ou exécuter le programme suspect au sein d'une machine virtuelle pour comparer son code à ceux de menaces déjà identifiées, ou son

[6] Xtended Detection and Response : Plateforme de collecte et de mise en corrélation des données générées par une infrastructure afin d'améliorer la visibilité des menaces à l'échelle de l'entreprise. Elle analyse, priorise et rationalise ces données avant de les envoyer aux équipes de sécurité dans un format normalisé via une console unique [53].
[7] Endpoint Detection and Response : Solution capturant toutes les activités des endpoints (terminaux informatiques) et s'appuyant sur des analyses avancées pour obtenir une visibilité en temps réel sur l'état des terminaux. L'EDR peut neutraliser une attaque en cours ou en limiter la propagation [53].
[8] Managed Detection and Response : Solution de sécurité des endpoints sous forme de service, cette solution intègre un EDR en plus de services complémentaires tels que la surveillance continue, le threat hunting ou encore l'intervention guidée [53].

comportement pour le comparer à des modèles de comportements potentiellement dangereux. D'un autre côté, l'analyse comportementale observe l'activité au niveau du système (pas uniquement du programme) et relève les actions qui paraissent malveillantes : requêtes vers un serveur inconnu, modifications de fichiers ou demandes d'accès à certains emplacements de la mémoire. Les deux méthodes sont complémentaires et peuvent cohabiter sur le même système [16].

3.2.2. Publiciels

Aussi connus sous le nom « adware », ce type de programme va diffuser des annonces sur l'écran de l'appareil infecté. Les données des utilisateurs peuvent ensuite être capturées pour être revendues à des annonceurs sans leur consentement. Il est à noter que ce type de programme peut aussi exister sous une forme « légitime » et complètement légale. Exemples notables : Fireball et Appearch [17].

3.2.3. Logiciels espions

Ces programmes malveillants se cachent sur l'appareil de la victime et effectuent de la surveillance d'activité ainsi que du vol d'informations. Leur propagation se fait via plusieurs vecteurs comme des chevaux de Troie, le regroupement avec un logiciel légitime ou encore l'exploitation de vulnérabilités de logiciels. Exemples notables : CoolWebSearch et Gator [17].

3.2.4. Rançongiciels et cryptovirus

Conçus dans l'unique but de recherche de profit, cette famille de programmes malveillants est développée pour empêcher les utilisateurs d'accéder à leur système et/ou leurs données tant qu'une rançon n'est pas versée. Cette rançon est généralement à payer en cryptomonnaie afin de limiter la traçabilité. Ce type d'attaque a pris une ampleur considérable depuis plusieurs années en s'attaquant à tous type d'entités, de la toute petite entreprise à la multinationale, en passant par les particuliers. La probabilité d'être visé par une attaque de type ransomware est extrêmement élevée par sa facilité de mise en place par l'attaquant et la rentabilité financière pour ce dernier. Exemples notables : WannaCry, Cryptolocker et Phobos [17].

3.2.5. Chevaux de Troie

Camouflé en logiciel légitime, le cheval de Troie peut être un point d'entrée de choix pour les pirates puisqu'il est possible ensuite d'y faire exécuter tout type de code pour exploiter l'appareil, récupérer des données ou effectuer toute action sur la machine infectée. L'utilisateur pensant ouvrir un programme légitime, il est aussi possible d'obtenir relativement facilement les privilèges d'administrateur. Exemples notables : Qbot et TrickBot [17].

3.2.6. Vers

Les vers sont des programmes malveillants assez courants, ils ont la particularité de pouvoir se reproduire de façon autonome pour infecter d'autres machines sans intervention extérieure. La charge utile transportée par un vers peut avoir divers objectifs comme le vol d'informations, l'installation de ransomware, la suppression de fichiers ou la création de botnets. Exemple notable : SQL Slammer [17].

3.2.7. Virus

Est défini comme virus un élément de code qui s'insère dans une application et s'exécute lors de son lancement. Un virus peut avoir des objectifs divers comme le vol de données, le lancement d'une attaque DDoS ou l'installation d'un ransomware. Il est généralement partagé par mail ou récupéré par téléchargement sur internet. Exemple notable : Stuxnet [17].

3.2.8. Enregistreurs de frappe

Aussi connus sous l'appellation « keylogger », cette famille de maliciels se rapproche de celle des logiciels espions. Ils enregistrent chaque appui sur une touche du clavier, cela peut être utilisé à des fins malveillantes pour voler des mots de passes, des informations bancaires ou d'autres informations confidentielles. Il est à noter que ces logiciels peuvent aussi être installés à des fins légitimes comme pour du contrôle parental ou le monitoring d'activité des employés d'une entreprise. Exemple notable : Snake Keylogger [17]

3.2.9. Bots et botnets

Un bot permet à un attaquant de contrôler à distance une machine, la force de l'attaque réside dans la propagation d'un bot afin de créer un botnet : un réseau d'ordinateurs zombies. Ils peuvent contaminer des millions d'appareils sans être détectés puis s'activer simultanément pur faire des dégâts importants. Ils sont particulièrement utilisés dans le cadre d'attaques DDoS

(Distributed Denial of Service), l'envoi de mails de phishing ou la diffusion d'un maliciel. Exemples notables : Andromeda et Marai [17].

3.2.10. Programmes malveillants sans fichier

Apparu en nombre depuis 2017, ce type d'attaque assez nouveau possède la particularité de n'écrire aucune donnée sur le disque de la machine attaquée, toute l'attaque est exécutée en mémoire vive. De ce fait l'attaque ne laisse aucune trace dès que la mémoire vive est effacée (au redémarrage par exemple). Leur empreinte faible et l'absence de fichier à analyser rend leur détection plus difficile par les antivirus traditionnels. Exemples notables : Frodo, Number of the Beast et The Dark Avenger [17].

3.2.11. Les nouvelles techniques de propagation

A la manière des logiciels légitimes, les maliciels sont aujourd'hui très évolués et maintenus via des mises à jour pour les faire évoluer. Deux exemples vont être brièvement présentés.

3.2.11.1. Rançongiciel BlackBasta

Détecté pour la première fois en février 2022, ce rançongiciel en C++ provoque un redémarrage de la machine en mode sans-échec. Cela peut sembler sans incidence particulière mais une partie des solutions de sécurité sont désactivées en mode sans-échec, le programme malveillant va donc être indétectable et pourra facilement chiffrer les fichiers stockés sur la machine.

A la suite d'une mise à jour, le rançongiciel est même capable de se connecter à l'Active Directory avec l'annuaire LDAP afin d'obtenir une liste des machines du réseau. Il se copie ensuite sur chacune d'entre-elle puis exécute son script malveillant [18].

3.2.11.2. Logiciel espion OnionPoison

Détecté pour la première fois en août 2022 alors qu'il était actif depuis le mois de janvier, ce programme imite le navigateur internet TOR, qui est bloqué en Chine. Ciblant majoritairement des utilisateurs chinois cherchant à se procurer illégalement ce navigateur offrant des fonctionnalités de pseudonymat, il était quasiment indétectable excepté en en analysant le code source.

Quelques éléments notables de ce logiciel malveillant sont l'absence de signature, une version de Mozilla Firefox modifiée qui ne se mettra pas à jour (afin d'éviter les correctifs de sécurité) et l'historique de navigation qui est stocké différemment sur le disque. Les DLLs

(Dynamic Link Libraries) du programme étants extrêmement similaires à ceux du programme légitime, il est très difficile de détecter sa présence sur le système. Une fois installé ce logiciel espion permet d'exécuter des commandes à distance sur le système, la récupération de l'historique de navigation et l'obtention des identifiants WeChat et QQ [18].

3.3. La Cyber Threat Intelligence : comprendre les menaces

3.3.1. Définition de la Cyber Threat Intelligence

La CTI est une discipline de la cybersécurité qui vise à rassembler, analyser et diffuser du renseignement sur les cybermenaces. Elle regroupe l'étude de marquants, de méthodes d'attaque utilisées par les pirates informatiques et de l'exploitation des vulnérabilités présentes dans les systèmes d'information actuels.

Cette discipline permet aux organisations de comprendre le paysage des menaces cybernétiques, d'identifier les vulnérabilités de leurs systèmes et ainsi mettre en place les mesures de sécurité adéquates. L'importance de la CTI est croissante à mesure que les cybermenaces évoluent et se complexifient.

3.3.2. Obtenir des données de Cyber Threat Intelligence

Les données de CTI sont variées et peuvent être de différents types :

- Des « observables » représentent des informations techniques pouvant être utiles pour des investigations (adresse IP, hash de fichier, url…).
- Des « marquants » ou « indicateurs de compromissions » sont des observables liés de façon avéré à une activité malveillante.
- Des rapports sur les groupes d'attaquants connus et leur mode opératoire (tactiques, techniques et procédures).
- Des rapports d'analyse sur des vulnérabilités critiques (CVEs).

Ces données peuvent provenir de différentes sources de réputations. Il existe des renseignements d'origine sources ouvertes dits OSINT (open source intelligence) qui peuvent représenter jusqu'à 80% des informations des services de renseignement [19]. En effet, dans un contexte actuel de big data, un grand nombre de données est accessible directement sur le web à qui sait les chercher correctement. Sans compter les nombreux forums de discussions qui traitent de ces sujets dont les articles sont parfois même écrits par des cybercriminels. On peut ici penser à la plateforme de messagerie Telegram détournée pour devenir une mine

d'informations pour les cybercriminels (et ainsi la cyberdéfense) [20]. De plus, il existe des bases de réputations payantes que les organisations peuvent se procurer afin d'enrichir leurs données de CTI. On peut citer pour exemple Palo Alto Network [21], ThreatConnect [22] ou encore Cybereason [23]. Enfin, les entreprises peuvent également se constituer leur propre banque de données de CTI des contextes métiers ou encore des localisations géographiques des sites de l'organisation. Ces renseignements peuvent provenir des systèmes de sécurité de l'organisation, des journaux d'audit et des équipes de réponse aux incidents.

A l'échelle nationale, le CERT-FR[9] possède également un rôle dans l'obtention de données de CTI pour les organisations. En effet, puisqu'il a pour mission de traiter les alertes et réagir aux attaques informatiques, l'organisme réalise de nombreuses analyses techniques d'attaques et récupère ainsi un grand nombre d'informations. De plus, le CERT-FR effectue une veille technologique permanente sur les produits majeurs du marché afin d'en détecter les vulnérabilités. Les publications récurrentes de l'organisme (alertes de sécurité, rapports menaces et incidents, indicateurs de compromission ou encore avis de sécurité) représentent une banque de ressource de CTI importante pour les organisations.[10]

3.3.3. Le cycle de vie des données de Cyber Threat Intelligence

Avant la CTI, une certaine "Roue de la douleur" était présente. La victime subissait une attaque, commandait une mise à jour de sécurité ou installait un nouveau produit curatif. Elle se faisait de nouveau attaquer et procédait encore une fois à une mise à jour et le cycle se répétait.

Aujourd'hui, la CTI permet d'acquérir et de développer une stratégie d'anticipation pertinente en se nourrissant des attaques passées. Pour ce faire, elle se base sur un cycle du renseignement en 6 étapes (voir *Figure 2*) :

1. Direction

   Les décideurs fixent ce sur quoi l'équipe de renseignement devrait se concentrer. C'est un état des lieux de la situation et une définition du besoin.

[9] Le CERT-FR est le CERT (Computer Emergency Response Team) ou CSIRT (Computer Security Incident Response Team) Gouvernemental et National Français et traite, sur le plan technique, les incidents de cybersécurité. Le CERT-FR est porté par la Sous-Direction Opérations de l'Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI) [58].

[10] Les sources de données de CTI citées dans cette partie du rapport ne représentent pas une liste exhaustive.

2. Collection

   La majorité des données est collectée via des ressources OSINT telles que l'analyse de réseaux sociaux ou encore des rapports des éditeurs de sécurité.

3. Processing

   Les données collectées peuvent être structurées ou non, il est donc nécessaire de les traiter, ce processus est automatisé, on peut distinguer 3 sous-étapes qui permettent de rendre les informations exploitables par les machines et les humains mais aussi d'apporter un degré de confiance dans chaque information :

   a. Normalisation
   b. Qualification
   c. Enrichissement

4. Analysis

   C'est une étape principalement manuelle et qui doit être réalisée par un humain, l'analyste vient apporter son expertise dans le but de générer des relations et interpréter le schéma global de l'attaque. L'objectif est aussi de contextualiser l'information.

5. Dissemination

   Une fois toutes les étapes d'analyse effectuées, la diffusion de l'information est cruciale. Il est à noter que la variété des acteurs ciblés (Risk Manager, SOC ou CERT) pose des contraintes sur le rendu des analyses qui doivent donc être compréhensibles par tous.

6. Feedback

   Cette dernière étape permet d'évaluer si la réponse aux besoins définis a été correctement effectuée mais aussi si le déroulement du processus a été efficace.

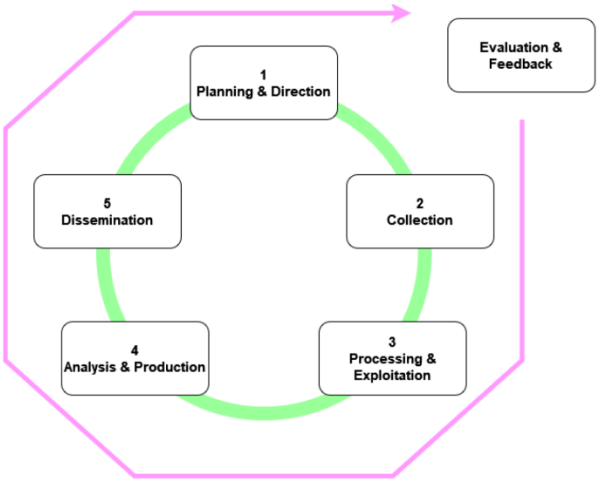

Figure 2 - Le processus de Cyber Threat Intelligence par Airbus Protect

Tout ce cycle permet de se poser les bonnes questions et de se renseigner de manière efficace sur ses adversaires dans le but de mieux les connaitre et ainsi construire et améliorer sa défense de manière efficace [24].

3.3.4. Le scoring des données de la CTI

Bien que la maîtrise de la Cyber Threat Intelligence représente un atout de taille pour les organisations, elle constitue également de nombreux défis. L'un d'entre eux est d'éviter la surcharge de données. Face à la complexification des méthodes d'attaques, la multiplication des attaquant et l'augmentation significative de la surface potentielle d'attaque, il devient de plus en plus difficile de répertorier l'ensemble données.

De plus, ces informations ont une durée de vie limitée en raison de la nature changeante du paysage des menaces cybernétiques. Les données sur les menaces peuvent rapidement devenir obsolètes à mesure que de nouvelles vulnérabilités sont découvertes, que de nouveaux logiciels malveillants sont développés et que les acteurs de la menace changent leurs tactiques, techniques et procédures (TTP).

Afin de tenir compte du caractère éphémère et changeant des données, ces dernières se voient souvent attribuer une date d'expiration après laquelle elles ne seront plus considérées comme valides, ainsi qu'un score. Le score de threat intelligence est une note attribuée à un indicateur de compromission ou à tout autre information sur les menaces qui indique la probabilité que l'information soit liée à une activité malveillante. Ces scores sont utiles en cas d'analyse afin d'aider les professionnels à prendre des décisions plus rapidement ou à prioriser les évènements de sécurité.

L'un des défis liés à la CTI est de faire vivre le score. Souvent il repose sur une combinaison de différents facteurs tels que la source de l'information, l'âge de l'information ou encore la gravité de la menace. De manière générale, les scores sont représentés sur une échelle de 1 à 100, 0 indiquant que l'information ne correspond probablement pas une activité malveillante et 100 indiquant que l'information correspond très fortement à une activité malveillante. Cependant, l'échelle spécifique utilisée peut varier d'une organisation à l'autre. Par exemple, Microsoft Defender Threat Intelligence pratique également le scoring de réputation pour les hôtes, domaines, ou adresses IPs. Les indicateurs de compromissions sont ainsi classés en différentes catégories : inconnu, neutre, suspect ou malveillant [25]. L'objectif des organisations est alors de mettre régulièrement le score des marquants à jour afin d'avoir une connaissance la plus exacte possible du paysage des cybermenaces.

3.3.5. L'utilisation de la CTI au sein d'un SOC

3.3.5.1. Les SOCs : centres névralgiques de la cyberdéfense

Un SOC (Security Operation Center ou centre des opérations de sécurité) est à la fois un processus et une équipe visant à surveiller et protéger l'ensemble d'une infrastructure informatique. Ce dernier peut être interne, externalisé ou hybride. Son enjeu principal est la détection d'évènements de cybersécurité en temps réel, indépendamment de la source, de l'heure de la journée ou du type d'attaque et y faire face le plus rapidement et le plus efficacement possible. Les SOC font face à de nombreux défis, notamment la détection de menaces encore méconnues cherchant à contourner les mécanismes de sécurité mis en place. Ils sont alors conçus pour s'adapter continuellement afin de pouvoir couvrir tout nouveau périmètre d'une organisation où des données stratégiques seraient à risque.

Les SOCs s'appuient grandement sur les données de la CTI pour une cybersécurité renforcée. Si le SOC se concentre sur la surveillance et la protection des systèmes d'information en temps réel, la CTI fournit les renseignements nécessaires pour anticiper les menaces et mieux les combattre. Ainsi, le SOC peut prioriser les incidents de sécurité en fonction des scores de threat intelligence des marquants, mettre en place des mesures de protection ciblées et enrichir les systèmes de détection d'intrusion avec des indicateurs de compromission connus.

3.3.5.2. Structurer les données de la CTI grâce au langage STIX[11]

STIX (Structured Threat Information Expression) est un langage normalisé et structuré conçu pour représenter les informations relatives aux menaces cyber. Ce standard a été développé en 2012 aux États-Unis par le CERT-US du Department of Homeland Security, puis repris par les organisations MITRE et OASIS[12].

La version 2 de STIX & TAXII (Trusted Automated eXchange of Intelligence Information), validée en 2017, a intégré le standard CybOX (Cyber Observable eXpression) et a progressivement abandonné le format XML au profit de JSON (JavaScript Object Notation). La version 2.1 de ces standards a récemment été validée, apportant des améliorations par rapport à la version précédente.

STIX définit 12 objets, appelés « STIX Domain Objects » (SDOs), qui permettent de catégoriser divers éléments d'informations liés aux menaces, tels que les vulnérabilités, les campagnes, les indicateurs, les kits d'intrusion, les schémas d'attaque et les logiciels malveillants. La combinaison de ces objets via un système de relations offre des représentations simples ou complexes de la CTI. En outre, STIX définit deux objets relationnels appelés les « STIX Relationship Objects » (SROs), à savoir l'observation et la relation, qui décrivent comment les objets SDOs sont liés les uns aux autres [26].

3.3.5.3. OpenCTI : une plateforme opensource de Cyber Threat Intelligence[13]

OpenCTI est une plateforme de renseignement sur les menaces conçue pour aider les organisations à collecter, gérer, analyser et partager des informations sur les menaces. Elle offre une multitude de fonctionnalités, notamment la centralisation des données de renseignement sur les menaces, la création de relations entre les indicateurs de compromission, l'analyse de l'activité malveillante, la génération de rapports et bien plus encore. L'orientation vers

[11] Toutes les informations présentes dans cette partie proviennent de la note d'information « STIX & TAXII : au cœur de la Threat Intelligence » rédigée le 21 octobre 2023 par Loïc Blondeau dans le cadre de la rédaction de la revue de littérature inhérente à ce projet (voir *Annexe 29*).

[12] Organization for the Advancement of Structured Information Standards : association professionnelle à but non lucratif. Sa mission est de développer, soutenir et promouvoir des projets de standardisation. C'est l'une des rares structures de normalisation reconnue par l'ISO (International Standards Organisation) pour développer des standards de type Publicly Available Specification.

[13] Les informations présentes dans cette partie proviennent de la note d'information « Note d'informations – SOC » rédigée le 10 octobre 2023 par Loïc Blondeau dans le cadre de la rédaction de la revue de littérature inhérente à ce projet.

OpenCTI a été motivée par plusieurs facteurs, le premier étant qu'il est open source, développé sur une initiative de l'ANSSI, en partenariat avec le CERT-EU. De ce fait, il est accessible gratuitement et bénéficie d'une communauté active de développeurs et d'utilisateurs qui contribuent à son amélioration continue. Aujourd'hui, l'entreprise gestionnaire de la plateforme est Filigran. OpenCTI a aussi été choisi du fait de sa popularité dans les entreprises de la cybersécurité.

OpenCTI se base sur la notion de « Knowledge graph » (« graphe de connaissances » en français). Les entités représentent les nœuds et les bords explicitent les relations entre les différentes entités. Ces bords possèdent des attributs et des propriétés. Pour expliciter plus clairement, les entités représentent les objets d'intérêt dans le domaine de la CTI, tels que les acteurs malveillants, les logiciels malveillants, les victimes d'attaques, les indicateurs de compromission (IOCs) ou encore les techniques et procédures utilisées par les attaquants [27].

L'objectif principal de la plateforme est de créer un outil permettant de capitaliser les informations techniques (telles que les marquants) et non techniques (telles que la victimologie par exemple) et de les relier à leur source primaire. Une fois que les données ont été capitalisées et traitées par les analystes d'OpenCTI, de nouvelles relations peuvent être déduites des relations. Cela permet à l'utilisateur d'extraire et d'exploiter des connaissances significatives à partir des données brutes. Ce modèle facilite alors grandement l'analyse avancée des données sur les menaces, permettant d'identifier des tendances, des modèles et des liens cachés entre les entités.

OpenCTI utilise le format STIX 2 pour représenter son modèle de données. STIX est un langage normalisé et structuré pour représenter les informations sur les cybermenaces, et en particulier des indicateurs de compromission. L'utilisation de STIX 2 assure la compatibilité d'OpenCTI avec d'autres outils de CTI, facilitant l'intégration et le partage de données [28].

La plateforme OpenCTI est une plateforme open source modulable. Elle bénéficie d'une communauté active de contributeurs. Ainsi, de nombreux connecteurs, plugins, clients et contenus sont créés par Filigran et sa communauté [29]. L'écosystème dynamique d'OpenCTI constitue un atout majeur pour la plateforme. En favorisant l'innovation, le partage des connaissances et l'adoption par un large éventail d'utilisateurs, l'écosystème contribue à faire d'OpenCTI un outil puissant pour la gestion du renseignement sur les menaces et la cybersécurité [29].

Plusieurs piliers composent l'écosystème de la plateforme :

- **La communauté OpenCTI** : Composée d'utilisateurs, de développeurs et de contributeurs, elle participe au développement de nouvelles fonctionnalités, à la correction de bugs et à la création de contenu.
- **Les intégrations tierces** : l'architecture ouverte de la plateforme permet son intégration avec divers outils de cybersécurité. Ces intégrations permettent d'étendre les fonctionnalités d'OpenCTI et de l'adapter à des besoins spécifiques.
- **Les partenaires technologiques** : De nombreux partenaires technologiques proposent des solutions et des services complémentaires à OpenCTI. Les connecteurs de partenaires support sont développés et entretenus par les fournisseurs d'origine.

3.4. L'Intelligence Artificielle (IA) : outil pour les cybermenaces ou atout pour la cyberdéfense.

3.4.1. Définition de l'Intelligence Artificielle

L'Intelligence Artificielle est un processus qui vise à imiter l'intelligence humaine au travers d'algorithmes informatiques [1]. Elle donne à la machine la capacité d'analyser et de prendre des décisions. Au sein de l'Intelligence Artificielle, on distingue généralement deux sous-disciplines basées sur l'apprentissage que sont le Machine Learning et le Deep Learning.

Le Machine Learning, ou apprentissage automatique, se base sur des algorithmes pour reproduire un comportement grâce à un grand nombre de données d'entrainement. Le Deep Learning, sous-discipline du Machine Learning, s'apparente encore davantage au fonctionnement du cerveau humain, en ajoutant des niveaux d'abstraction à l'aide de couches de neurones virtuels. Il peut comprendre des concepts avec plus de précision en fonction du nombre neurones qu'il possède [1].

En Machine Learning, on considère généralement deux types d'apprentissages distincts : l'apprentissage supervisé et l'apprentissage non supervisé. Dans l'apprentissage supervisé, les données sont étiquetées en amont dans le dataset avant d'être utilisées par l'algorithme. A l'inverse, dans l'apprentissage non-supervisé, les données ne sont pas étiquetées. L'avantage de l'apprentissage supervisé est qu'il est généralement plus efficace car le modèle va être capable d'apprendre les relations qui existent entre les données d'entrée et celles de sortie. En revanche, c'est souvent difficile de créer des datasets étiquetés car cela prend du temps. D'un

autre côté, l'apprentissage non supervisé peut être utilisé pour déterminer des tendances, et trouver des motifs qui se répètent dans les données d'entrée.

L'apprentissage supervisé inclut des algorithmes destinés à résoudre des problèmes de classification et de régression. Dans la classification, on cherche à attribuer des étiquettes à des données d'entrée après qu'un algorithme se soit entrainé sur des données étiquetées à l'avance. Par exemple, on pourrait s'en servir pour distinguer des courriels « spam » et « non-spam ». La régression, quant à elle, consiste à prédire des valeurs continues, comme l'estimation du prix d'une maison en fonction de ses caractéristiques. D'autre part, l'apprentissage non supervisé regroupe des algorithmes qui traitent, par exemple, des problèmes de clustering et d'association. Le clustering vise à regrouper des données en différents ensembles similaires sans que ces données n'aient été étiquetées au préalable. Il peut, entre autres, être utile pour regrouper des clients en segments de marché similaires. L'association identifie des relations ou des motifs intéressants dans de grandes bases de données et peut notamment permettre de déterminer quels produits sont souvent achetés ensemble dans un magasin.

Ces diverses branches de l'Intelligence Artificielle permettent de résoudre un large éventail de problèmes dans de nombreux domaines d'application. Parmi ces domaines, la cybersécurité se distingue comme un enjeu majeur.

3.4.2. L'IA : un potential outil pour les cybermenaces

L'avènement de l'accès facile et souvent gratuit à des intelligences artificielles a permis aux attaquants un perfectionnement de leurs méthodes déjà existantes. En effet le phishing et la personnification de ce dernier sont devenus très efficaces grâce à la génération rapide et suffisamment crédible de texte ou même de médias tels que des images, vidéos, articles ou messages [30]. L'Intelligence Artificielle est aussi facilement capable de générer un mail de spam personnalisé destiné à une personne en utilisant simplement quelques informations que cette dernière aura laissé traîner sur internet. Toujours dans la catégorie du phishing, l'Intelligence Artificielle peut être utilisée également pour générer de faux site web à but malveillant (récupération d'identifiants, coordonnées bancaires ou tout autre type de donnée personnelle). Ces nouvelles technologies sont aussi capables d'usurper l'identité de personnes à la base de photos de la personne (deepfake). Cette pratique s'est accentuée pendant la pandémie du Covid-19 où les cybercriminels ont eu accès à de nombreuses données biométriques grâces aux visioconférences [31].

L'Intelligence Artificielle a aussi rendu beaucoup plus accessible le codage et le réseau, facilitant la création de scripts malveillants à envoyer par mail. On peut citer ChatGPT (OpenAI) et Codex (OpenAI) qui permettent d'aider à l'écriture de tels outils en quelques minutes. On arrive donc à des cyberattaques qui sont presque automatisées étant donné l'élaboration accélérée de ces types d'attaques.

3.4.3. L'IA : un atout pour la cyberdéfense

L'Intelligence Artificielle apporte des avantages non-négligeables pour une cybersécurité plus efficace. Tout d'abord, sa capacité à traiter un important volume de données sept jours sur sept et 24 heures sur 24 a l'avantage d'améliorer considérablement les aptitudes de détection. Elle permet également de décharger les humains de certaines tâches chronophages comme de l'aide à l'identification d'activités inhabituelles, la détection d'activités anormales ou la prévention de l'exploitation de failles.

Dans le cadre de la réponse à incident, l'IA peut également prendre un rôle important et permettre un travail plus efficace. En effet les plans de restauration après attaque comportent de nombreuses tâches pouvant être automatisées comme la réinstallation des solutions de sécurité sur les machines, la vérification des clés de registre ou la modification des règles dans les pares-feux. On peut également noter que les tests d'intrusion ont la possibilité d'être accélérés et automatisés davantage grâce à l'Intelligence Artificielle.

L'IA ne diminue donc pas le besoin d'expertise une fois les informations collectées mais ses capacités permettent à une structure d'être plus efficace dans sa globalité. On peut aussi noter sa puissance dans le domaine de la gestion de la donnée afin de cartographier, référencer, sauvegarder ou même chiffrer des grandes quantités de données [32].

4. Partie recherche sur l'IA, la CTI et la cybersécurité

4.1. Revue de littérature

La première étape du projet a consisté en une phase de documentation sur les thèmes gravitant autour de notre sujet de projet. A ce titre, plusieurs articles ont été étudiés et des fiches de lecture associées ont été rédigées (voir *9.3* et *9.4)*.

La méthode de travail adoptée a d'abord consisté en la documentation sur 5 thèmes essentiels pour la bonne compréhension du contexte : CTI, IA, SOC & IOC, CKC et Sandbox.

Dès lors que ces concepts de bases ont été bien cernés, des notions plus avancées et plus spécifiques ont été étudiées.

Afin d'avancer sur la revue de littérature et d'en assurer sa bonne qualité, des critères de sélection des articles (voir *Tableau 2*) ont été définis.

Tableau 2 - Critères de sélection des articles

Critères de sélection des articles	Date d'écriture supérieure à 2019
	DOI reconnu
	Mots clés : CTI, IA, Malware detection
	Données de comparaison détaillées

De plus, comme expliqué précédemment, un template pour les fiches de lecture a été élaboré afin de garantir la meilleure analyse possible et l'homogénéité du travail. Chaque article a été synthétisé à l'aide des éléments suivants :

- Titre ;
- Entête avec auteur(s), type de document, date de publication, lecteur(s), date de lecture, site/ouvrage et lien ;
- Problématique ;
- Résumé ;
- Moyens employés ;
- Solution trouvée ;
- Critique de la solution ;
- Remarques complémentaires.

La revue de littérature inhérente à ce projet est constituée de 14 fiches de lectures[14] (voir *Tableau 3*).

Tableau 3 - Liste des articles de la revue de littérature

Titre	Auteur(s)	Type	Date de publication	DOI

[14] Les critères de sélection ayant été établis au cours du projet, certains documents présents dans la revue de littérature peuvent dater d'avant 2019.

Cyber Threat Intelligence – Issue and Challenges	Md Sahrom Abu ; Siti Rahayu Selamat ; Aswami Ariffin ; Robiah Yusof.	Article	01/04/2018	DOI:10.11591/ijeecs.v10.i1.pp371-379
Artificial Intelligence in Cyber Threat Intelligence	Roumen Trifonov ; Ognyan Nakov ; Valderi Mladenov.	Article	06/01/2019	DOI:10.1109/ICONIC.2018.8601235
Application Security through Sandbox Virtualization	Liberios Vokorokos ; Anton Baláž ; Branislav Madoš.	Article	01/2015	N/A
Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity	Sherali Zeadally ; Erwin Adi ; Zubair Baig ; Imran A. Khan.	Article	20/01/2020	DOI:10.1109/ACCESS.2020.2968045
Avast-CTU Public CAPE Dataset	Branislav Bošanský ; Dominik Kouba ; Ondˇrej Manhal ; Thorsten Sick ; Viliam Lisý ; Jakub Kˇroustek ; Petr Somol.	Article	06/09/2022	DOI:10.48550/arXiv.2209.03188
Fashion Images Classification using Machine Learning, Deep Learning and Transfer Learning Models	Bougareche Samia ; Zehani Soraya ; Mimi Malika.	Article	03/06/2022	DOI: 10.1109/ISPA54004.2022.9786364
2021 SANS Cyber Threat Intelligence (CTI) Survey	Rebekah Brown ; Robert M. Lee.	Livre Blanc	18/01/2021	N/A
The Unified Kill Chain Raising resilience against advanced cyberattacks	Paul Pols.	Livre Blanc	02/2023	N/A

A Comprehensive Survey on Identification of Malware Types and Malware Classification Using Machine Learning Techniques	Nagababu Pachhala ; S. Jothilakshmi ; Bhanu Prakash Battula.	Article	12/11/2021	DOI:10.1109/ICOSEC51865.2021.9591763
Algorithmes d'Intelligence Artificielle en Cybersécurité & Intégration en environnements contraints	Stéphane Morucci ; Stéphane Davy ; Nicolas Raux ; Jérémy Scion ; Guillaume Lerouge ; Marc Le Nué ; Nathan Rydin ; Dorian Screm.	Article	11/2019	N/A
Artificial intelligence in cyber security: research advances, challenges, and opportunities	Zhimin Zhang ; Huansheng Ning ; Feifei Shi ; Fadi Farha ; Yang Xu ; Jiabo Xu ; Fan Zhang ; Kim Kwang Raymond Choo.	Article	10/03/2021	DOI:10.1007/s10462-021-09976-0
CyTIME - Cyber Threat Intelligence ManagEment framework for automatically generating security rules	Eunsoo Kim ; Kuyju Kim ; Dongsoon Shin ; Beomjin Jin ; Hyoungshick Kim.	Article	20/06/2018	DOI:10.1145/3226052.3226056
Machine Learning for Security and the Internet of Things: The Good, the Bad, and the Ugly	F. Liang ; W. G. Hatcher ; W. Liao ; W. Gao ; W. Yu.	Article	22/10/2019	DOI:10.1109/access.2019.2948912

Performance Evaluation of Machine Learning Classifiers in Malware Detection	Umesh V. Nikam ; Vaishali M. Deshmuh.	Article	13/06/2022	DOI:10.1109/ICDCECE53908.2022.9793102

4.2. Analyse critique de l'existant

Les recherches effectuées ont montré que l'intégration de l'Intelligence Artificielle et de la CTI aux outils de détections des cybermenaces est un sujet d'actualité qui a déjà vu plusieurs implémentations concrètes dans des entreprises [33]. Le livre blanc « 2021 SANS CTI Survey » de Rebekah Brown et Robert M. Lee [34] (voir *Annexe 17*) fait l'étude des tendances sur l'utilisation de la CTI dans 300 entreprises. Son utilisation est croissante et est majoritairement dédiée au processing (voir *3.3.3*) des données. Les articles « CTI – Issue and Challenges » [35] (voir *Annexe 11*) et « CyTIME - CTI ManagEment framework for automatically generating security rules » [36] publiés en 2018 cherchent à définir et standardiser les données de la CTI (voir *Annexe 22*). En quelques années cette discipline s'est précisée et a évoluée. Désormais la définition est claire (voir *3.3.1*) et le protocole le plus utilisé afin de standardiser les données est STIX (voir *3.3.5.2*).

L'apprentissage automatique, quant à lui, est utilisé afin d'identifier et des classifier les logiciels malveillants. Cependant dans « A Comprehensive Survey on Identification of Malware Types and Malware Classification Using Machine Learning Techniques » (voir *Annexe 19*) écrit par Nagababu Pachhala, S. Jothilakshmi et Bhanu Prakash Battula dans le cadre de l'ICOSEC (International Conference on Smart Electronics and Communication) [37], il est expliqué que la mise en œuvre de ce type de solution est complexe pour les entreprises de petite taille car il leur est difficile de collecter un grand nombre de données. L'article « Performance Evaluation of Machine Learning Classifiers in Malware Detection » de Umesh V. Nikam et Vaishali M. Deshmuh [38] (voir *Annexe 24*) étudie la performance de dix algorithmes d'apprentissage automatique dans le cadre de la classification de malware. C'est l'algorithme XGBOOST qui se démarque de cette étude. Cet algorithme sera utilisé dans le cadre de ce projet cependant le résultat est à nuancer puisque l'article ne se concentre que sur dix algorithmes et ne s'intéresse qu'au système d'exploitation Android.

De plus, l'article « Algorithmes d'Intelligence Artificielle en Cybersécurité & Intégration en environnements contraints » écrit par Stéphane Morucci, Stéphane Davy, Nicolas Raux *et*

al. [39] (voir *Annexe 20*) souligne l'importance de la pertinence des données et met en lumière l'utilité de la séparation de l'apprentissage et du scoring des marquants dans le cadre des environnements contraints. Sur le même sujet, l'article « Avast-CTU Public CAPE Datasets » (voir *Annexe 15*) écrit par Branislav Bošanský, Dominik Kouba, Ondˇrej Manhal *et al.* [40] souligne le caractère éphémère des données pouvant rendre les datasets d'entrainement rapidement obsolètes. Pour faire face aux erreurs de prédictions et à l'obsolescence des données Zhimin Zhang, Huansheng Ning, Feifei Shi *et al.* proposent d'intégrer un facteur humain dans la boucle grâce à son modèle « Human in the Loop » théorisé dans l'article « Artificial intelligence in cyber security: research advances, challenges, and opportunities » [41] (voir *Annexe 21*). Bien qu'intéressant, ce type de modèle est plus complexe ce qui peut rendre son déploiement et sa maintenance plus difficile.

Dans le document « Machine Learning for Security and the Internet of Things: The Good, the Bad, and the Ugly » (voir *Annexe 23*), les auteurs (F. Liang, W. G. Hatcher, W. Liao *et al.*) avertissent sur la vulnérabilité des modèles d'IA qui peuvent être utilisés de manière malveillante. Si un modèle d'IA capable de détecter les menaces est corrompu (par un maliciel par exemple), il devient caduc.

4.3. Question de recherche et proposition de réponse

L'élaboration de la revue de lecture a permis de définir une problématique qui se décline en une question de recherche. Le corpus documentaire expose l'existence de modèles combinant IA et détection de malware. Néanmoins, ces solutions sont faillibles et ne s'adaptent pas à tout type d'entreprise. De plus, la gestion de données CTI, souvent internes, représente un défi important. L'IA, la CTI et malheureusement les cybermenaces se développant à un rythme effréné, les systèmes de détection doivent évoluer parallèlement. C'est cette problématique qu'adresse notre question de recherche : Comment améliorer les techniques d'apprentissage automatique existantes pour détecter des malwares connus et inconnus en intégrant des données de la CTI ?

Au vu des documents étudiés, la réponse à la problématique doit prendre en compte un certain nombre d'éléments critiques du corpus documentaire. Afin de solutionner la nécessité d'une quantité importante de données soulignée dans l'article « A Comprehensive Survey on Identification of Malware Types and Malware Classification Using Machine Learning Techniques » [37], la solution doit intégrer un lien fort avec une plateforme collaborative de CTI telle que OpenCTI. De même, le modèle « Human in the Loop » [41] semble intéressant à

implémenter, mais cela nécessite de partir d'un modèle déjà entraîné pour limiter l'action humaine au début de l'implémentation. La duplication des algorithmes de reconnaissance peut également être une solution afin d'améliorer la fiabilité des résultats. Le lien avec une plateforme de données de la CTI combinée à l'expertise d'un analyste permettrait l'enrichissement continu du modèle, résolvant ainsi le problème d'obsolescence des données soulevé par l'article « Avast-CTU Public CAPE Dataset » [40].

La partie réalisation de ce projet représente les premières étapes du développement de la solution. Tout d'abord, la création d'un environnement de test de détonation de malware et de récupération de logs constitue un prérequis à la solution imaginée. Puis, l'implémentation et la comparaison des modèles les plus pertinents de la revue de littérature a donné lieu à la conception de plusieurs modèles de classification des malwares.

5. Réalisation et mise en œuvre

5.1. Environnement de tests et récupération de logs

5.1.1. Outils utilisés

5.1.1.1. La sandbox CAPEv2

Pour la sandbox de la partie réalisation, le choix de l'équipe s'est orienté vers CAPEv2 car contrairement aux autres sandbox (Fireye malaware analyse plug in, Valkyrye Comodo ou sandboxie par exemple), CAPEv2 est gratuit et open source. En plus d'être simple d'utilisation tout en ayant des fonctionnalités avancées pour l'analyse de malware, CAPEv2 permet une intégration directe à Open-CTI. CAPEv2 est un « tout en un » regroupant plusieurs applications : l'Intrusion Detection System (IDS) Suricata, la sandbox Cuckoo et l'hyperviseur KVM ce qui rend l'installation et l'utilisation de la solution plus simple. Le choix de l'équipe a aussi été motivé par l'existence de datasets de logs d'analyse de logiciels par CAPEv2 qui permettent à l'équipe d'entrainer les modèles d'Intelligence Artificielle avec un grand nombre de donnée en même temps qu'elle développe l'environnement de test CAPEv2.

CAPEv2 Sandbox est un logiciel permettant l'analyse de fichiers suspects. Pour ce faire, il utilise des composants personnalisés qui surveillent le comportement des processus malveillants lorsqu'ils s'exécutent dans un environnement isolé. Ce choix s'appuie sur une série de critères techniques et fonctionnels qui répondent à des besoins spécifiques.

Pour commencer, de nombreux datasets s'appuient sur les logs générés par CAPEv2. Ces datasets sont indispensables à l'entraînement de notre modèle d'Intelligence Artificielle. La mise en place de cet outil générera le bon format en entrée pour notre modèle. Il offre une analyse approfondie des malwares, y compris la capture des interactions réseau, des modifications du système de fichiers et des comportements en mémoire. Il fournit des rapports détaillés et structurés, facilitant l'interprétation des résultats.

Pour optimiser l'utilisation de CAPEv2, plusieurs dépendances clés sont nécessaires. Chacune de ces dépendances ont été choisies pour ses caractéristiques spécifiques qui complètent les fonctionnalités de CAPEv2.

Le système d'exploitation utilisé est Ubuntu. Les performances ainsi que la compatibilité avec de nombreux outils de sécurité et de virtualisation en fait un choix évident pour le déploiement de CAPEv2 qui est optimisé pour Ubuntu.

L'outil de virtualisation KVM (Kernel-based Virtual Machine) est utilisé. KVM est une solution de paravirtualisation natif au noyau Linux, permettant de créer et gérer des machines virtuelles au travers de l'interface VirtualManager. Il offre des performances proches du matériel, ce qui s'avère crucial pour faire tourner un système d'exploitation Windows en parallèle de la machine Ubuntu.

Comme cité précédemment, CAPEv2 est une sandbox. Cela signifie qu'au travers d'une machine virtuelle isolée, un code logiciel potentiellement dangereux peut s'exécuter sans affecter les ressources du réseau ou les applications locales. Pour cela, CAPEv2 fait appel à Cuckoo Sandbox.

Enfin, CAPEv2 fait également appel à l'IDS (Intrusion Detection System) Suricata. Cela permet une analyse en temps réel du trafic réseau sur la machine virtuelle.

Un résumé du fonctionne de CAPEv2 est disponible en *Figure 3*, provenant de la documentation officielle de CAPE Sandbox [42].

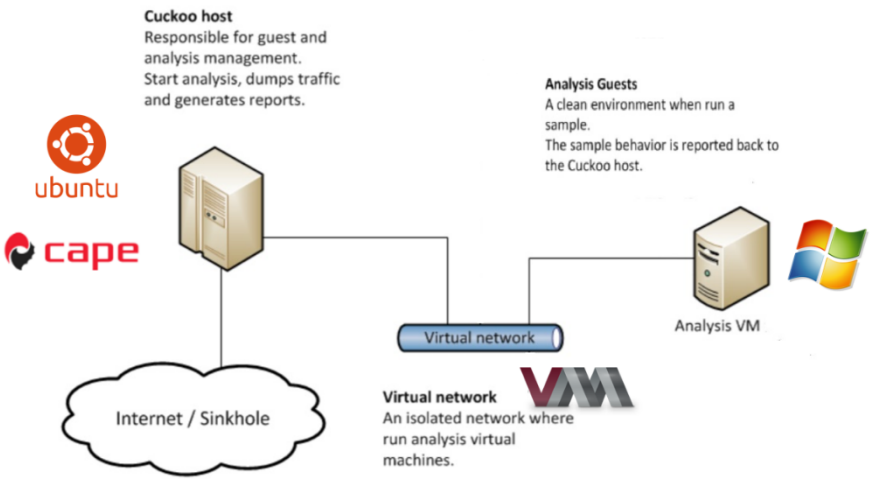

Figure 3 – Architecture du réseau d'environnement

5.1.2. Installation et configuration

Pour l'installation de l'environnement de test, l'équipe a opté pour une machine Ubuntu car cape est un logiciel tournant sur Ubuntu et qui nécessite une machine virtuelle (VM) Windows. Le choix a été fait d'installer une machine virtuelle Windows 7 car il existe énormément de vulnérabilités connues sur cet OS (Operating System). Au début l'équipe avait fait une première installation sur une machine Windows avec une VM Ubuntu et une VM Windows dans la VM Ubuntu. Au vu des problèmes de performances rencontrés avec cette installation, il a été décidé de procurer une machine physique pour mener à bien cette mission.

Le projet de CAPEv2 met à disposition deux scripts d'installation [43]. Un premier nommé "cape2.sh" pour installer le programme et un deuxième "kvm-qemu.sh" pour installer l'hyperviseur. Ensuite, il faut configurer les fichiers de configuration de CAPEv2 en fonction des besoins. Tous les fichiers sont stockés dans le répertoire "/opt/CAPEv2/conf/". Enfin, il est possible d'accéder à l'interface graphique de CAPEv2 sur l'adresse local sur le port 8000.

Pour les détails plus techniques de l'installation, une documentation sur l'installation de CAPEv2 a été rédigée pour le projet et est disponible en annexe.

5.1.3. Mise en œuvre

Pour démarrer une analyse, il faut se rendre dans l'onglet « Submit » de l'interface graphique et d'importer le fichier à analyser (voir *Figure 4*) :

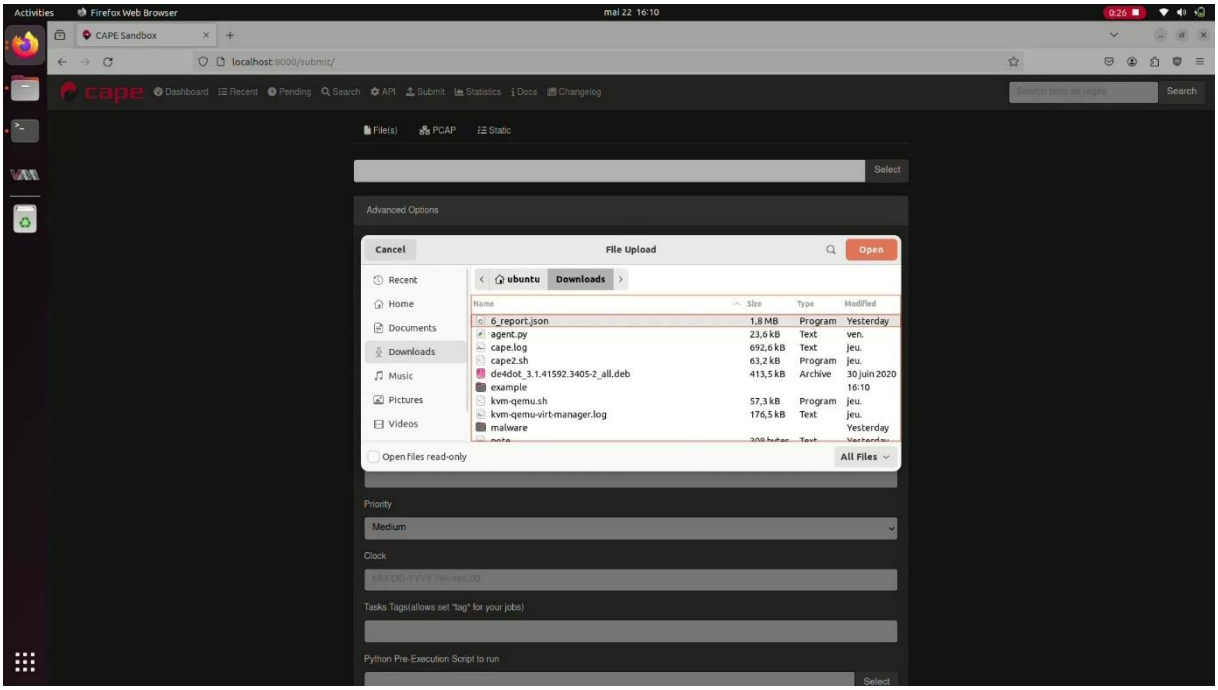

Figure 4 - Page d'importation du fichier à analyser

Plusieurs options sont ensuite disponibles pour analyser le fichier, comme le ”timeout” qui permet de préciser la durée de l'analyse. Il y a aussi des options pour autoriser l'accès de la machine virtuelle au réseau ou non, des commandes python à exécuter avant, ou encore des options pour définir s'il faut récupérer les fichiers d'analyses réseau ou une copie de la mémoire vive pour des analyses poussées après exécution. Une fois que les paramètres désirés ont été sélectionnés, il faut cliquer sur le bouton pour lancer l'analyse et ensuite démarrer la machine virtuelle via l'hyperviseur. La machine virtuelle s'éteindra toute seule à la fin de l'analyse. [44]

5.1.4. Résultats

Afin de pouvoir visualiser l'interface et les résultats générés par CAPEv2, le logiciel nous propose plusieurs écrans détaillés :

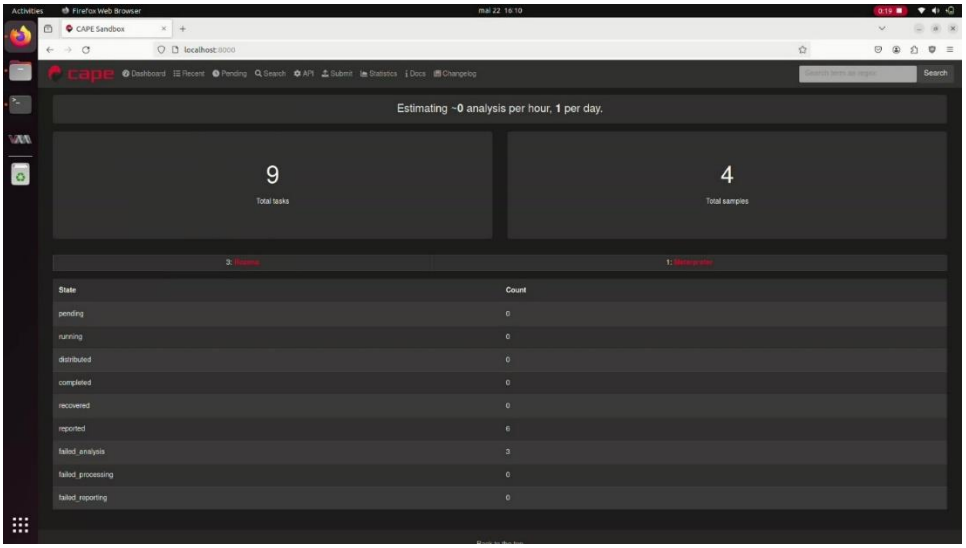

Figure 5 - Page d'accueil de CAPEv2

Cette page présente les analyses antérieures et en cours (voir *Figure 5* ci-dessus). Elle met en évidence le nombre d'analyse en fonction de leur état. Notamment, elle indique s'il existe des analyses en cours ou en attente, ce qui peut parfois expliquer pourquoi une analyse ne s'exécute pas directement.

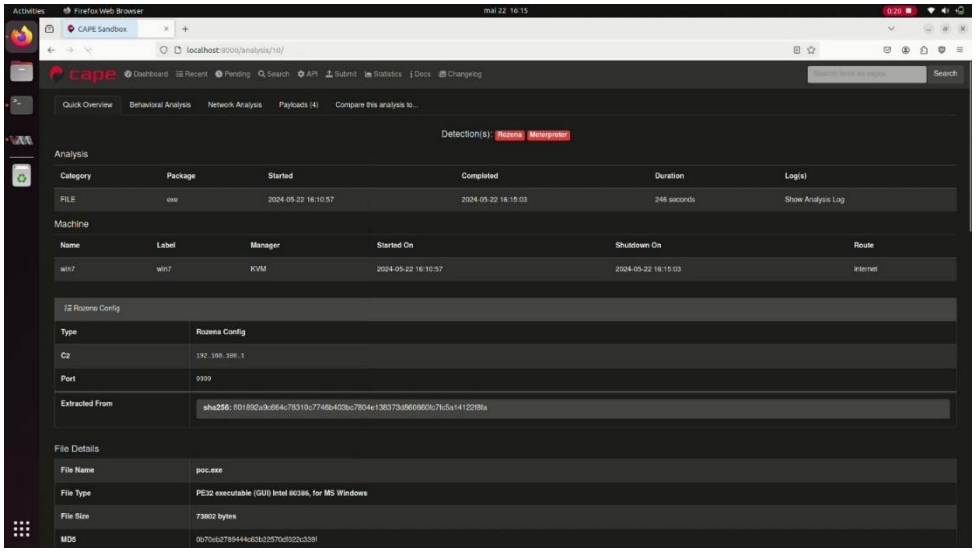

Figure 6 - Tableau de bord présentant le compte rendu d'une analyse

La page de compte rendue d'analyse présente toutes les informations recueillies durant l'analyse (voir *Figure 6* - Tableau de bord présentant le compte rendu d'une analyse ci-dessus). En rouge, dans « detection(s) », le type de malware estimé est affiché. La page présente les métadonnées de l'analyse comme la date, la durée, la route réseau utilisé… La signature du malware est donnée, cela est utile pour constituer une base de données de logiciels dangereux dans le but de récolter des donnés de Cyber Threat Intelligence. Des détails supplémentaires sont affichés pour justifier l'interprétation des résultats comme étant un certain type de malware.

Sur les autres onglets, il est possible de trouver plus de détails, notamment sur le comportement chronologique du logiciel comme dans l'onglet « Behavior Analysis ». A l'instar de l'onglet « Network Analysis » qui permet de faire la même chose mais au niveau des paquets réseaux. Cet onglet est très intéressant car il intègre l'IDS Suricata, ce qui permet de voir si quelque chose a été détecté au niveau du réseau. Il est également possible de comparer deux analyses.

Le fichier en format « json » présentant tous les résultats au format CAPEv2 est disponible au bas de cette page.

5.1.5. Intégration de openCTI à CapeSandbox

Comme dit précédemment, CAPEv2 permet une intégration à OpenCTI grâce à la fonctionnalité *Cape Connector* développée par l'équipe d'OpenCTI [44]. Cette intégration facilite la collecte, l'analyse et la mise en relation des données sur les menaces en temps réel. CAPEv2 peut donc envoyer des rapports détaillés d'analyse de malwares à OpenCTI. L'automatisation de ce flux de données assure ainsi une mise à jour continue de la base de données d'OpenCTI. De plus, cette synergie entre CAPEv2 et OpenCTI optimise les processus de réponse aux incidents en fournissant des indicateurs de compromission (IoCs) précis et exploitables qui permettent d'améliorer la résistance globale des systèmes de sécurité.

5.2. Modèles d'IA

5.2.1. Outils utilisés

Les modèles d'IA ont été créés sur des notebooks en langage Python. En effet, Python a été choisi car c'est l'un des langages de code les plus utilisés dans le domaine de l'IA, et ce pour plusieurs raisons. D'abord, il possède un écosystème de bibliothèques variées ce qui permet d'utiliser des implémentations d'algorithmes optimisées et efficaces et de ne pas avoir à tout développer à partir de zéro. Il permet également de visualiser facilement les données traitées et les résultats obtenus, et il est simple à comprendre même pour les novices en programmation.

La bibliothèque principale utilisée pour tous les traitements relatifs au machine learning est Scikit-learn [40]. C'est une bibliothèque très robuste qui inclut tous les éléments utiles à la préparation des données, à la création des modèles, à leur entrainement, à leur évaluation et à leur utilisation finale. Scikit-learn permet également de régler un grand nombre de paramètres pour adapter les modèles à des cas d'utilisation précis.

Pour la partie deep learning, la bibliothèque choisie est TensorFlow [45]. Elle est open-source et développée par Google, ce qui la rend solide et avec une API simple à utiliser. Elle possède les mêmes avantages que Scikit-learn pour le deep learning. TensorFlow a été choisi par rapport à PyTorch, qui est une autre bibliothèque Python permettant de faire du deep learning, car TensorFlow est plus adapté pour des projets à grande envergure et pour créer des modèles évolutifs.

5.2.2. Le dataset

Le dataset provient d'une collaboration entre Avast Software et le centre d'IA de l'université polytechnique de Prague [40]. Il est constitué de 48 976 échantillons de malware, classés dans les dix familles de malwares (voir *Tableau 4*).

Tableau 4 - Familles de malwares du dataset utilisé

Adload	**Emotet**	**HarHar**	**Lokibot**	**njRAT**	**Qakbot**	**Swisyn**	**Trickbot**	**Ursnif**	**Zeus**
704	14 429	655	4 191	3 372	4 895	12 591	4 202	1 343	2 594

Chaque famille de malware est associée à un des six types suivants : « banker », « trojan », « pws », « coinminer », « rat » et « keylogger ». Les échantillons ont, pour la plupart, été collectés au cours des années 2017 jusqu'à 2019, ce qui permet de voir les évolutions des changements dans chaque famille au cours du temps.

Pour chaque échantillon, il y a deux types de données qui sont récoltées. Les données comportementales contiennent un résumé de l'analyse du comportement du malware comme les fichiers accédés, les clés de registres, les mutexes et les appels API. Les données statiques contiennent toutes les propriétés statiques du logiciel, comme sa signature, sa somme de contrôle ou son point d'entrée.

5.2.3. La méthodologie CRISP-DM

Pour concevoir efficacement une solution d'IA, il est important d'avoir une méthodologie bien définie. Dans le cadre de ce projet, la méthodologie suivie se nomme CRISP-DM, ou Cross-Industry Standard Process for Data Mining [46]. Elle permet d'avoir un aperçu du cycle de vie de l'exploration de données (voir *Figure 7*).

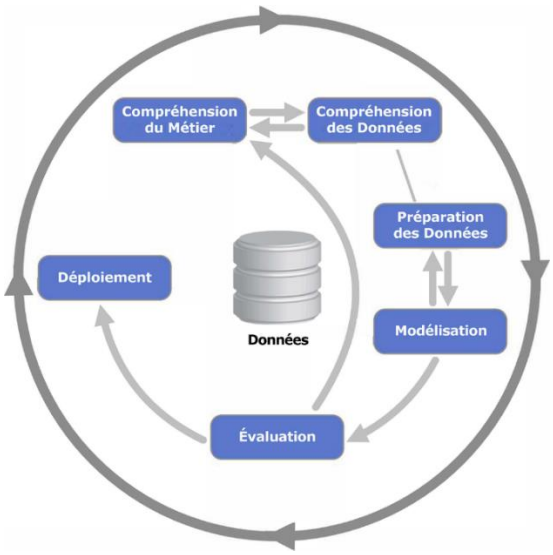

Figure 7 - Diagramme du processus CRISP-DM

Le processus commence par une phase de compréhension du métier. Avant toute manipulation des données, il est impératif de bien cerner les besoins métier. La phase de recherche du projet a permis d'obtenir une vue d'ensemble sur les attentes en termes d'apprentissage automatique appliqué à la CTI. La seconde phase du processus est la compréhension des données. Une fois le jeu de données rassemblé, il est crucial de comprendre sa structure et ses caractéristiques pour pouvoir les exploiter au mieux et ainsi créer des modèles performants. Ensuite, les données doivent être préparées, c'est-à-dire traitées pour optimiser les performances des modèles. En parallèle, la modélisation peut commencer : un premier modèle est créé, entrainé avec les données préparées, puis adapté pour obtenir de bons résultats. L'avant-dernière étape est l'évaluation. Il faut regarder comment le modèle fonctionne et vérifier qu'il répond aux attentes du métier. Si ce n'est pas le cas, les étapes précédentes peuvent être répétées pour continuer à l'améliorer. Enfin, le déploiement est envisagé. Cela consiste à mettre en place l'environnement complet qui sera utilisé par le métier dans des situations réelles.

5.2.4. Les différents algorithmes utilisés

Dans le cadre de ce projet, les algorithmes qui ont été étudiés sont des algorithmes de classification entrainés de manière supervisée. Leur but est de déterminer la classe à laquelle appartient une donnée, à partir de ses caractéristiques qui ont été étudiées pendant la phase d'entrainement. En effet, dans le cadre de la détection de malware, il est intéressant de savoir à quelle famille de malware appartient un logiciel donné. En ayant un dataset suffisamment fourni d'exemples de malwares étiquetés, les algorithmes vont être capables de déterminer, par exemple, si un malware fait partie de la famille des Trickbot ou des Emotet. Il existe un large éventail d'algorithmes différents, dont les plus connus sont comparés ici (voir *Tableau 5*) :

Tableau 5 - Comparaison des algorithmes de machine learning

Algorithme	Précision	Vitesse	Evolutivité	Robustesse	Données
Decision Trees	Peut capturer des relations complexes non linéaires dans les données	Relativement rapide	Peut gérer de grandes quantités de données et de caractéristiques	Bonne performance sur des données nouvelles	Fonctionne bien avec de petites ou moyennes quantités de données étiquetées

Random Forest	Peut capturer des relations complexes non linéaires dans les données, souvent plus précis qu'un seul arbre de décision	Relativement rapide à entrainer, mais peut être lent à classifier	Peut gérer de grandes quantités de données et de caractéristiques	Bonne performance sur des données nouvelles	Fonctionne bien avec de petites ou moyennes quantités de données étiquetées
Gradient Boosting (XGBoost)	Peut capturer des relations complexes non linéaires dans les données, souvent plus précis qu'un seul arbre de décision	Relativement lent à entraîner, mais rapide à classifier	Peut gérer de grandes quantités de données et de caractéristiques	Bonne performance sur des données nouvelles	Fonctionne bien avec des quantités moyennes ou grandes de données étiquetées
Naive Bayes	Performe bien lorsque l'hypothèse d'indépendance tient	Très rapide	Peut gérer de grandes quantités de données et de caractéristiques	Bonne performance sur des données nouvelles	Fonctionne bien avec de petites quantités de données étiquetées
K-Nearest Neighbors	Peut capturer des relations complexes non linéaires dans les données	Relativement lent à entraîner, mais rapide à classifier	Peut gérer de grandes quantités de données, mais peut être lent pour des espaces de caractéristiques de haute dimension	Bonne performance sur des données nouvelles	Fonctionne bien avec des quantités moyennes ou grandes de données étiquetées
Support Vector Machine	Peut capturer des relations complexes non linéaires dans les données	Relativement lent à entraîner, mais rapide à classifier	Peut gérer de grandes quantités de données, mais peut être lent pour des espaces de caractéristiques de haute dimension	Bonne performance sur des données nouvelles	Fonctionne bien avec des quantités moyennes ou grandes de données étiquetées
Artificial Neural Network	Peut capturer des relations complexes non linéaires dans les données	Relativement rapide à entraîner, mais peut être lent à classifier	Peut gérer de grandes quantités de données et de caractéristiques	Bonne performance sur des données nouvelles	Fonctionne bien avec de grandes quantités de données étiquetées

Le choix des algorithmes utilisés dans cette étude repose donc sur des raisons bien définies. Tout d'abord, K-Nearest Neighbor a été sélectionné pour sa simplicité, sa flexibilité et ses performances même sur des ensembles de données de taille moyenne. Random Forest a été choisi en raison de son faible risque de surapprentissage et de sa capacité à estimer l'importance des différentes caractéristiques. XGBoost a été privilégié pour son efficacité à gérer les déséquilibres dans les classes. Enfin, les réseaux de neurones artificiels ont été intégrés pour leur aptitude à comprendre des relations complexes, leur approche personnalisable et leur performance sur des données volumineuses. La diversité de ces algorithmes garantit une complémentarité, exploitant les forces spécifiques de chacun pour améliorer la classification des malwares.

5.2.4.1. K plus proches voisins

Le premier algorithme est l'algorithme des k plus proches voisins, aussi connu sous le nom de KNN [47]. C'est une approche de classification supervisée qui regroupe les données en fonction de leur proximité par rapport à leurs plus proches voisins. Les étapes importantes de l'algorithme sont :

1. Pour chaque nouvelle instance à classer, on calcule sa distance dans l'espace des caractéristiques par rapport à toutes les instances du dataset d'entrainement.
2. Les *k* instances les plus proches (c'est-à-dire avec les distances les plus petites) sont sélectionnées.
3. Un vote majoritaire est effectué parmi les *k* voisins sélectionnés et la classe résultante est attribuée à la nouvelle instance.

5.2.4.2. Random Forest

Le second algorithme étudié est le Random Forest [48]. C'est une technique d'apprentissage basée sur de multiples arbres de décision entrainés sur des sous-ensembles de données légèrement différents. Un arbre de décision correspond à un ensemble de questions qui portent sur les caractéristiques des données et organisées sous forme de branches. Chacune de ces branches amène vers un nœud final, appelé nœud de feuille, qui correspond à une classe du dataset (voir *Figure 8*).

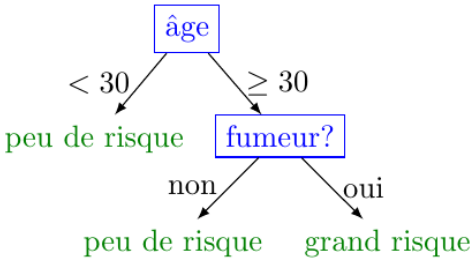

Figure 8 - Arbre de décision sur le risque d'avoir un accident cardiovasculaire

En combinant plusieurs arbres décisionnels, on effectue un vote majoritaire sur tous leurs résultats pour déterminer la prédiction finale de notre modèle (voir *Figure 9*).

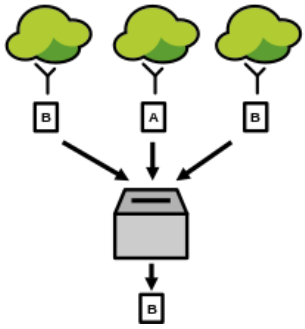

Figure 9 - Vote majoritaire effectué à la suite du calcul des prédictions intermédiaires

5.2.4.3. Extreme Gradiant Boosting

Un autre algorithme qui a été étudié est appelé Extreme Gradiant Boosting, ou XGBoost [49]. C'est une technique d'apprentissage supervisé basée sur le principe des arbres de décision boostés. Contrairement à la méthode Random Forest qui entrainent les arbres en parallèle, XGBoost entraine des arbres séquentiellement, chaque nouvel arbre corrigeant les erreurs du précédent (voir *Figure 10*). Le processus de boosting peut être décrit en quatre étapes :

1. On commence avec un arbre de décision simple qui fait une première prédiction.
2. On calcule les résidus (les erreurs de prédiction) du premier arbre.
3. Un nouvel arbre est entrainé pour prédire spécifiquement ces résidus.
4. Les prédictions de tous les arbres sont combinées pour donner la prédiction finale. Chaque arbre apporte une petite correction, rendant le modèle de plus en plus précis.

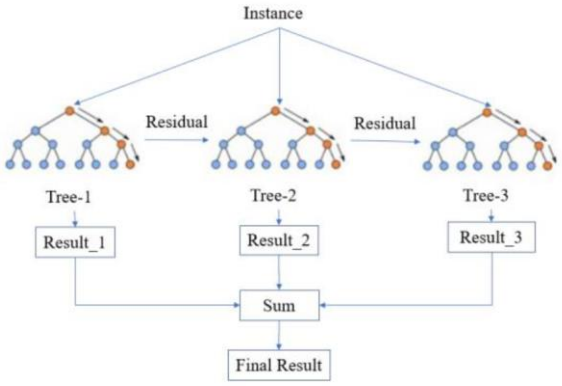

Figure 10 - Schéma d'un XGBoost

5.2.4.4. Réseau de neurones artificiels

Enfin, le dernier algorithme est le réseau de neurones artificiels, ou Artificial Neural Network (ANN) en anglais [50]. Contrairement aux algorithmes précédemment explicités, l'Artificial Neural Network (ANN)[15] est un algorithme de Deep Learning. Il est composé de plusieurs couches de nœuds : une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Chaque nœud est connecté à tous les nœuds des couches adjacentes, et possède son propre poids et seuil. Si la sortie d'un nœud est supérieure à son seuil, alors le nœud sera activé et il enverra de l'information aux nœuds de la couche suivante, exactement comme un neurone biologique. Plus le réseau de neurones est conséquent, plus l'algorithme sera capable d'extraire des caractéristiques abstraites des données pour prédire la classe à laquelle elles appartiennent (voir *Figure 11*).

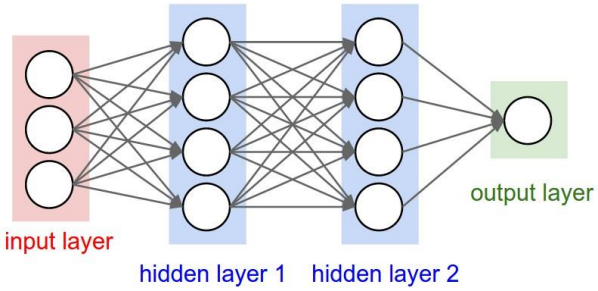

Figure 11- Schéma d'un réseau de neurones artificiels

[15] Réseau de Neurones Artificiels en français.

5.2.5. Mise en œuvre

5.2.5.1. Prétraitement du dataset

Avant de créer des modèles et de les entrainer sur le dataset, il est important de préparer les données pour qu'elles correspondent bien à ce qu'attendent les algorithmes.

Le premier traitement effectué est la sélection des caractéristiques importantes du jeu de données. Chaque échantillon du dataset possède énormément de caractéristiques, et certaines sont bien plus pertinentes que d'autres quand il s'agit de les classer dans des familles de malware. Par exemple, il a été déterminé que la donnée « icon_hash » n'apportait pas beaucoup d'information utile au modèle, alors que la donnée « executed_commands » était primordiale. Scikit-learn nous permet de visualiser l'importance de ces caractéristiques et de choisir lesquels seront gardés pour la suite. Voici la liste des 13 caractéristiques utilisées pour ce projet, classées par ordre d'importance (voir *Figure 12*) :

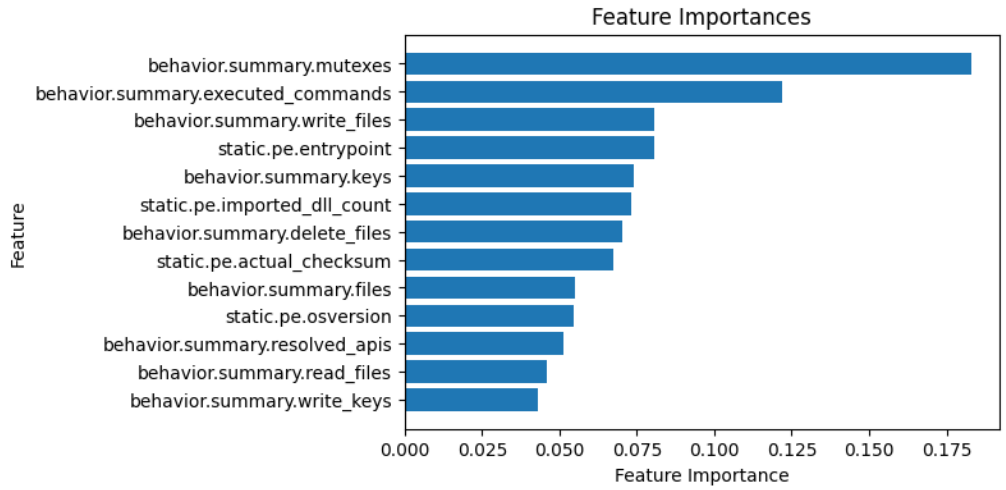

Figure 12 - Importance des caractéristiques du dataset

Un deuxième traitement important à effectuer sur les données est d'encoder toutes les données textuelles en données numériques, pour que les algorithmes puissent les comparer et effectuer des calculs. Pour ça, Scikit-learn propose la classe « LabelEncoder » qui va créer un dictionnaire avec tous les mots présents dans nos données et associé à chaque mot un nombre, puis il va transformer chaque mot en sa représentation numérique.

Un autre traitement important est la normalisation des données. En effet, cela augmente généralement les performances et la stabilité des modèles, notamment ceux basés sur les calculs de distance comme l'algorithme des K plus proches voisins.

Pour terminer, il faut également diviser les données en deux parties distinctes, de manière aléatoire, pour obtenir des données d'entrainement (80%) et des données d'évaluation (20%). Cela permet aux modèles créés d'être évalués sur des échantillons qu'ils n'avaient jamais rencontrés pendant l'entrainement, ce qui simule au mieux un cas réel et permet d'obtenir des résultats cohérents avec la réalité.

5.2.5.2. Implémentation des modèles

Chacun des algorithmes utilisés dans ce projet possède ses propres spécificités d'implémentation. Pour réussir à obtenir les meilleurs modèles possibles, il a été nécessaire de régler les paramètres mis à disposition par les bibliothèques Python pour observer leur influence sur les résultats et trouver les paramétrages optimaux.

Pour commencer, le modèle des K plus proches voisins a été implémenté grâce à la classe « KNeighborsClassifier » disponible dans la bibliothèque Scikit-learn. Le modèle a été créé en utilisant une fonction de pondération basée sur la distance des voisins. Autrement dit, plus un voisin est proche, plus sa classe aura d'importance dans le vote de détermination de la classe finale. De plus, le nombre de voisins sélectionnés pour déterminer la classe finale est $k = 5$.

Pour le modèle basé sur l'algorithme Random Forest et celui basé sur l'algorithme XGBoost, ce sont les classes « RandomForestClassifier » de la bibliothèque Scikit-learn et « XGBClassifier » de la bibliothèque « XGBoost » qui ont été utilisées. Les modèles ont été créés avec $n = 100$ arbres décisionnels différents, chacun utilisant $q = \sqrt{p}$ attributs, avec q le nombre d'attributs retenus dans chaque arbre et p le nombre total d'attributs de notre dataset.

Enfin, le réseau de neurones artificiels a été implémenté à l'aide de la classe « Sequential » disponible avec la bibliothèque TensorFlow. Le réseau est composé d'une couche d'entrée avec 13 nœuds, chacun représentant une caractéristique du dataset. Il possède quatre couches cachées avec respectivement 256, 128, 64 et 32 nœuds. Pour terminer, il est composé d'une couche de sortie avec 10 nœuds, chacun représentant une probabilité d'appartenance à une famille de malware. Lors de l'entrainement, il a été décidé d'utiliser la technique de dropout, ou décrochage, pour supprimer aléatoirement 30% des nœuds de chaque couche à chaque étape, et ainsi éviter le surapprentissage.

5.2.6. Résultats

Pour évaluer les performances et comparer les modèles d'IA créés, il est important d'utiliser plusieurs indicateurs clé de performance, ou KPI. Ces métriques sont calculées à partir des

prédictions réalisées par le modèle sur le jeu de données réservé à l'évaluation. Avant de comprendre comment sont calculés ces indicateurs, il est important de connaître la différence entre vrai positif, faux négatif et faux positif (voir *Figure 13*).

		Reality		
	Confusion matrix	**Class A**	**Class B**	**Class C**
Prediction	**Class A**	**True Positive A : TPA**	**False Negative B : FNB** **False Positive A : FPA**	**False Negative C : FNC** **False Positive A : FPA**
	Class B	**False Positive B : FPB** **False Negative A : FNA**	**True Positive B : TPB**	**False Negative C : FNB** **False Positive B : FPB**
	Class C	**False Positive C : FPC** **False Negative A : FPA**	**False Positive C : FPC** **False Negative B : FNB**	**True Positive C : TPC**

Figure 13 - Matrice de confusion multi-classe théorique pour 3 classes

Ce tableau montre que lorsque le modèle prédit correctement une classe, cela correspond à un vrai positif. En revanche, si le modèle prédit la classe « B » alors que la classe réelle est « A », c'est un cas de faux positif « B » et de faux négatif « A ». Un faux positif signifie que le modèle a prédit incorrectement une instance comme positive, et un faux négatif signifie qu'il a manqué une instance positive réelle.

Le premier indicateur clé utilisé pour mesurer la performance des modèles est l'accuracy. Elle mesure simplement la proportion de prédictions correctes parmi l'ensemble des prédictions effectuées. Une autre métrique présentée dans les résultats est la précision. Elle met en évidence la capacité du modèle à éviter les faux positifs. Pour chaque classe i, on peut la calculer de cette manière :

$$précision_i = \frac{\text{nb de documents correctements attribués à la classe } i}{\text{nb de documents attribués à la classe } i}$$

Parallèlement, le recall, ou rappel, évalue la capacité du modèle à éviter les faux négatifs et se calcule comme suit :

$$rappel_i = \frac{\text{nb de documents correctements attribués à la classe } i}{\text{nb de documents appartenant à la classe } i}$$

Pour finir, le F1-score, quant à lui, correspond à la moyenne harmonique entre la précision et le rappel. Il permet de fournir une vue d'ensemble de la performance du modèle et est calculé comme ça :

$$F_1 = 2 \cdot \frac{(précision \cdot rappel)}{(précision + rappel)}$$

Pour chaque modèle, les résultats sont présentés sous forme de tableau dans lequel on retrouve la précision, le rappel et le F1-score pour chaque famille de maliciel. De plus, chaque tableau présente l'accuracy du modèle, ainsi que la précision moyenne, le rappel moyen et le F1-score moyen, pondéré et non-pondéré (voir *Figure 14*, *Figure 15*, *Figure 16* et *Figure 17*).

	precision	recall	f1-score
Adload	1.00	0.90	0.95
Emotet	0.89	0.97	0.93
HarHar	1.00	1.00	1.00
Lokibot	0.94	0.94	0.94
Qakbot	1.00	0.98	0.99
Swisyn	1.00	0.98	0.99
Trickbot	0.95	0.91	0.93
Ursnif	1.00	0.75	0.86
Zeus	1.00	0.80	0.89
njRAT	0.91	0.92	0.91
accuracy			0.95
macro avg	0.97	0.92	0.94
weighted avg	0.95	0.95	0.95

Figure 14 - Tableau des résultats du KNN

	precision	recall	f1-score
Adload	1.00	1.00	1.00
Emotet	0.97	1.00	0.98
HarHar	1.00	1.00	1.00
Lokibot	0.97	0.97	0.97
Qakbot	1.00	0.96	0.98
Swisyn	1.00	1.00	1.00
Trickbot	0.98	0.95	0.96
Ursnif	0.93	0.72	0.81
Zeus	0.91	0.75	0.82
njRAT	0.95	0.98	0.96
accuracy			0.97
macro avg	0.97	0.93	0.95
weighted avg	0.97	0.97	0.97

Figure 15 - Tableau des résultats du Random Forest

	precision	recall	f1-score
Adload	1.00	1.00	1.00
Emotet	0.98	0.98	0.98
HarHar	1.00	1.00	1.00
Lokibot	0.95	0.96	0.96
Qakbot	1.00	0.98	0.99
Swisyn	1.00	1.00	1.00
Trickbot	0.92	0.87	0.89
Ursnif	1.00	0.71	0.83
Zeus	0.89	0.89	0.89
njRAT	0.94	0.97	0.95
accuracy			0.97
macro avg	0.97	0.93	0.95
weighted avg	0.97	0.97	0.97

Figure 16 - Tableau des résultats du XGboost

	precision	recall	f1-score
Adload	0.92	0.80	0.86
Emotet	0.91	0.93	0.92
HarHar	0.99	1.00	1.00
Lokibot	0.94	0.91	0.92
Qakbot	0.96	0.83	0.89
Swisyn	1.00	1.00	1.00
Trickbot	0.80	0.87	0.84
Ursnif	0.67	0.80	0.73
Zeus	0.84	0.55	0.67
njRAT	0.89	0.95	0.92
accuracy			0.93
macro avg	0.89	0.86	0.87
weighted avg	0.93	0.93	0.93

Figure 17 - Tableau des résultats de l'ANN

Il est également possible de visualiser l'évolution des performances d'un modèle au fur et à mesure de son entrainement. Plus le modèle s'entraine, plus son accuracy augmente (voir *Figure 18*). Cela permet notamment de détecter à partir de combien d'époques[16] le modèle arrête de s'améliorer.

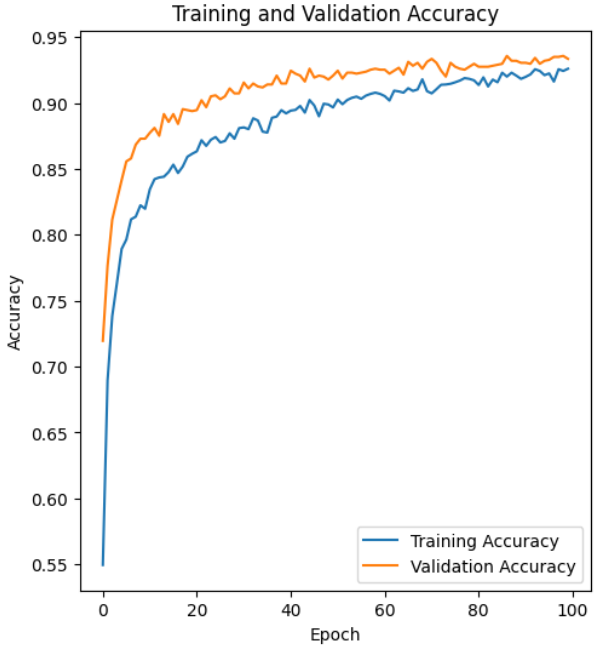

Figure 18 - Évolution de la précision au cours de l'entrainement de l'ANN

[16] Une époque en machine learning désigne un passage complet du jeu de données d'entrainement par l'algorithme. Autrement dit, après 100 époques, le modèle s'est entrainé 100 fois sur la totalité du dataset.

Les résultats permettent d'identifier les modèles les plus performants dans le cadre du projet. Avec une accuracy et un F1-score de 97%, ce sont les modèles Random Forest et XGBoost qui sont les plus adaptés. Le modèle KNN obtient une accuracy de 95% et un F1-score de 94%, et le modèle d'ANN possède une accuracy de 93% et un F1-score de seulement 87%. Cela peut s'expliquer par le fait que les modèles de deep learning nécessitent généralement un plus grand nombre de données d'entrainement pour discerner les caractéristiques de chaque classe, et qu'il faudrait encore améliorer l'architecture du modèle.

Un axe d'amélioration de ces modèles se situe dans l'ajout de logiciels légitimes au sein du dataset. Cela permettrait aux modèles de pouvoir faire la distinction entre un logiciel légitime et un malware, et ainsi éviter que les modèles prédisent une famille de malware dans le cas d'un logiciel légitime. Il est également important de préciser que les données d'entrainement proviennent de malwares détectés entre les années 2017 et 2019. Les méthodes de cyberattaques étant susceptibles d'évoluer rapidement, il est important de mettre à jour ce dataset dans le futur lorsque ces méthodes auront suffisamment évolué pour ne plus être détectées.

Ces modèles, couplés à l'environnement CAPEv2, offrent une solution complète de détection de malware. A partir de n'importe quel malware, il est possible de déterminer assez précisément la famille de malware auquel il appartient. En revanche, il ne faut pas se reposer uniquement sur les prédictions des modèles et il faut idéalement garder l'humain au centre de ce système.

6. Conclusion générale

Les objectifs de ce projet étaient d'élaborer une question de recherche claire et de proposer une solution innovante pour renforcer la posture de cybersécurité des organisations en exploitant les avancées actuelles en CTI et en IA.

La revue de lecture rédigée est composée de 14 fiches de lectures détaillées. L'analyse critique de l'existant a donné lieu à une problématique qui peut s'illustrer par la question de recherche suivante : Comment améliorer les techniques d'apprentissage automatique existantes pour détecter des malwares connus et inconnus en intégrant des données de la CTI ?

Ce travail a permis à l'équipe d'acquérir un grand nombre de connaissances sur différents sujets, à effectuer un travail de recherche scientifique précis et documenté et à synthétiser, communiquer, analyser et dépasser les conclusions des articles étudiés afin de proposer une solution innovante.

La solution proposée s'articule autour de plusieurs axes principaux : la création d'un système de détection complet, la mise en place de multiples modèles basés sur des algorithmes de classification différents et la prise en compte de l'intelligence humaine dans le processus décisionnel. Le résultat de la réalisation se compose d'un environnement de logs Ubuntu et d'un algorithme de ML de prédiction de famille de malware basé sur le modèle Random Forest. Le choix de ce modèle fait suite à la comparaison de différents algorithmes. L'environnement de logs basé sur CAPEv2 permet également la génération de rapport de données CTI qui peuvent être partager sur des plateformes collaboratives telles que OpenCTI. L'impossibilité d'obtenir un dataset de taille plus importante comprenant des logiciels légitimes dans le temps imparti, empêche le modèle de faire la distinction entre logiciel légitime et malveillant et restreint également la diversité des familles détectées. Il cherchera alors à classifier tout type de comportement comme appartenant à une famille de malware.

Ces résultats sont le fruit d'un travail de groupe efficace et d'une méthodologie de travail adaptative et appliquée sur l'ensemble de la durée du projet. Elle a ainsi permis de faire face aux défis rencontrés. L'équipe a su s'adapter aux changements d'attentes et ainsi transposer les recherches effectuées en une solution technique concrète. Etant donné l'ampleur du sujet, la revue de lecture rédigée pourrait être étoffée afin de suivre l'évolution des tendances et découvertes actuelles. Elle représente néanmoins la première approche d'un travail de recherche scientifique et d'une introduction à l'exercice de la thèse.

Ce projet présente de nombreuses opportunités et perspectives. Le modèle élaboré pourrait être déployé et intégré à une solution telle qu'un IDS. L'utilisation de Joblilb[17] permettrai l'exportation du modèle et donc son utilisation pratique. Les rapports de CAPEv2 et les conclusions du modèle peuvent constituer des données CTI supplémentaires. Les deux algorithmes qui se démarquent de la comparaison des modèles sont le Random Forest et le XGBoost, ils pourraient alors être utilisés en parallèle pour effectuer une reconnaissance avancée dans la solution finale. Enfin, la rédaction d'un article scientifique présentant les résultats de ce projet permettrait la contribution de l'équipe à la recherche en cybersécurité.

[17] Joblib est une bibliothèque Python qui permet d'exporter des modèles et autres données en tant que fichier sur le disque dur et de les importer ultérieurement.

7. Références

[1] Oracle, «Intelligence Artificielle, Machine Learning, Deep Learning : une histoire de poupées russes,» [En ligne]. Available: https://www.oracle.com/fr/database/deep-learning-machine-learning-intelligence-artificielle/. [Accès le 22 Mai 2024].

[2] Microsoft, «Copilot pour Microsoft 365,» [En ligne]. Available: https://www.microsoft.com/fr-fr/microsoft-365/business/copilot-for-microsoft-365.

[3] NIST, «National Vulnerability Database,» [En ligne]. Available: https://nvd.nist.gov/. [Accès le 16 Mai 2024].

[4] Forum InCyber, «SAVE THE DATE Forum InCyber Europe – 26, 27, 28 mars 2024,» [En ligne]. Available: https://incyber.org/evenement/save-the-date-forum-incyber-europe-26-27-28-mars-2024/.

[5] CNIL, «Cybersécurité,» [En ligne]. Available: https://www.cnil.fr/fr/technologies/cybersecurite. [Accès le 27 Mai 2024].

[6] ANSSI, «Panorama de la cybermenace 2023,» 27 Février 2024. [En ligne]. Available: https://www.cert.ssi.gouv.fr/cti/CERTFR-2024-CTI-001/. [Accès le 15 Mars 2024].

[7] NIST, «Glossary,» [En ligne]. Available: https://csrc.nist.gov/glossary/term/malware. [Accès le 17 mai 2024].

[8] Scrum.org, «What is Scrum?,» [En ligne]. Available: https://www.scrum.org/learning-series/what-is-scrum/what-is-scrum. [Accès le 21 Mai 2024].

[9] Asana, «Qu'est-ce que la méthode Kanban? Définition et outils,» 13 Février 2024. [En ligne]. Available: https://asana.com/fr/resources/what-is-kanban. [Accès le 21 Mai 2024].

[10] Équipe de projet, «Gestion de projet,» [En ligne]. Available: https://lunar-grin-c1d.notion.site/Projet-M2-Cyber-Threat-Intelligence-ef4ba242e5dd4d489a0f5a098cf34e15?pvs=4.

[11] IBM, «Présentation générale de CRISP-DM,» 17 Août 2021. [En ligne]. Available: https://www.ibm.com/docs/fr/spss-modeler/saas?topic=dm-crisp-help-overview. [Accès le 21 Mai 2024].

[12] Lockheed Martin, «Who we are,» [En ligne]. Available: https://www.lockheedmartin.com/en-us/who-we-are.html. [Accès le 10 Octobre 2023].

[13] LOCKHEED MARTIN , «GAINING THE ADVANTAGE, Applying Cyber Kill Chain Methodology to Network Defense».

[14] MITRE, «Who we are,» [En ligne]. Available: https://www.mitre.org/who-we-are. [Accès le 22 Mai 2024].

[15] P. Pols, «The Unified Kill Chain, Raising resilience against advanced cyber attacks,» 2023.

[16] Clubic, «Comment un antivirus détecte-t-il un virus ?,» 29 Mars 2024. [En ligne]. Available: https://www.clubic.com/antivirus-securite-informatique/logiciel-antivirus/article-853946-1-analyse-comportementale-comment-antivirus-veille-systeme.html. [Accès le 21 Mai 2024].

[17] kaspersky Labs, «Quels sont les différents types de programmes malveillants ?,» [En ligne]. Available: https://www.kaspersky.fr/resource-center/threats/types-of-malware. [Accès le 21 Mai 2024].

[18] SecureList, «Uncommon infection and malware propagation methods,» 05 Octobre 2022. [En ligne]. Available: https://securelist.com/uncommon-infection-and-malware-propagation-methods/107640/. [Accès le 22 Mai 2024].

[19] O. Le Deuff, L'Open Source Intelligence (OSINT) : origine, définitions et portée, entre convergence professionnelle et accessibilité à l'information dans I2D - Information, données & documents 2021/1 (n° 1), pages 14 à 20, 2021.

[20] L. Moalosi, «How Telegram is Becoming a Hub for Cybercrime!,» 2023.

[21] Palo Alto Network, «AutoFocus threat intelligence architecture,» [En ligne]. Available: https://www.paloaltonetworks.com/cortex/autofocus.

[22] ThreatConnect , «Unified Threat Library,» [En ligne]. Available: https://threatconnect.com/solutions/unified-threat-library/. [Accès le 22 Mai 2024].

[23] Cybereason, «Threat Intelligence,» [En ligne]. Available: https://www.cybereason.com/platform/threat-intelligence. [Accès le 22 Mai 2024].

[24] A. Khalifi, «OpenCTI, un incontournable dans l'analyse de la cybermenace,» 25 Juin 2020. [En ligne]. Available: https://www.capfi.fr/la-newsroom/blog/open-cti-un-incontournable-dans-l-analyse-de-la-cybermenace/. [Accès le 15 Mai 2024].

[25] Microsoft Defender, «Reputation scoring,» 20 Mai 2024. [En ligne]. Available: https://learn.microsoft.com/en-us/defender/threat-intelligence/reputation-scoring. [Accès le 22 Mai 2024].

[26] Orange Cyberdéfense, «STIX & TAXII : au cœur de la Threat Intelligence,» 05 Janvier 2021. [En ligne]. Available: https://www.orangecyberdefense.com/fr/insights/blog/securite-du-mail/stix-taxii-au-coeur-de-la-threat-intelligence. [Accès le 15 Mai 2024].

[27] S. H. e. Lhorus6, «Data model,» Juin 2023. [En ligne]. Available: https://docs.opencti.io/latest/usage/data-model/. [Accès le 15 Mai 2024].

[28] Filigran, «Introduction to the OpenCTI platform,» 19 Septembre 2022. [En ligne]. Available: https://youtu.be/2tBDnZYwmBs?si=waYv6V2GRbo0t0aH. [Accès le 15 Mai 2024].

[29] Filigran, «OpenCTI Ecosystem,» [En ligne]. Available: https://filigran.notion.site/OpenCTI-Ecosystem-868329e9fb734fca89692b2ed6087e76. [Accès le 15 Mai 2024].

[30] Université d'Oxford, «The Malicious Use of Artificial Intelligence:Forecasting, Prevention, and Mitigation,» 2018.

[31] K. B. C. E. M. N. M. S. Carta, «Video Injection Attacks on Remote Digital Identity Verification Solution Using Face Recognition,» chez *Proceedings of the 13th International Multi-Conference on Complexity, Informatics and Cybernetics: IMCIC 2022, Vol. II*, 2022.

[32] «Intelligence artificielle et cybersécurité : risques ou opportunités ?,» [En ligne]. Available: https://www.groupeonepoint.com/fr/nos-publications/intelligence-artificielle-et-cybersecurite-risques-ou-opportunites/. [Accès le 23 Mai 2024].

[33] L. Delabarre, «Comment l'IA a révolutionné le SOC ?,» 28 Novembre 2023. [En ligne]. Available: https://www.itforbusiness.fr/comment-l-ia-a-revolutionne-le-soc-70145. [Accès le 2024 Janvier 10].

[34] R. B. e. R. M. Lee, «2021 SANS Cyber Threat Intelligence (CTI) Survey,» 18 Janvier 2021.

[35] S. R. S. A. A. e. R. Y. Md Sahrom Abu, «Cyber Threat Intelligence – Issue and Challenges,» 1 Avril 2018.

[36] K. K. D. S. e. a. Eunsoo Kim, «CyTIME - Cyber Threat Intelligence ManagEment framework for automatically generating security rules,» 20 Juin 2018.

[37] S. J. e. B. P. B. Nagababu Pachhala, «A Comprehensive Survey on Identification of Malware Types and Malware Classification Using Machine Learning Techniques,» 2021.

[38] U. V. N. e. V. M. Deshmuh, «Performance Evaluation of Machine Learning Classifiers in Malware Detection,» 13 Juin 2022.

[39] S. D. N. R. e. a. Stéphane Morucci, «Algorithmes d'Intelligence Artificielle en Cybersécurité & Intégration en environnements contraints,» 2019.

[40] B. Bosansky, D. Kouba, O. Manhal, T. Sick, V. Lisy, J. Kroustek et P. Somol, «Avast-CTU Public CAPE Dataset,» 2022.

[41] H. N. F. S. e. a. Zhimin Zhang, «Artificial intelligence in cyber security: research advances, challenges, and opportunities,» 10 Mars 2021.

[42] «What is CAPE?,» CAPE Sandbox, [En ligne]. Available: https://capev2.readthedocs.io/en/latest/introduction/what.html. [Accès le 22 Mai 2024].

[43] Cape , «Cape Sandbox Book,» [En ligne]. Available: https://capev2.readthedocs.io/en/latest/. [Accès le 13 05 2024].

[44] CalamityZ, "CTI installation & connection," [Online]. Available: https://github.com/calamityz/CTI-Installation-Connection/blob/main/CTI%20Installation%20%26%20Connection.md#-CAPE-%E2%86%94-OpenCTI. [Accessed 27 05 2024].

[45] TensorFlow, «TensorFlow,» Google, [En ligne]. Available: https://www.tensorflow.org. [Accès le 24 Mai 2024].

[46] «Présentation générale de CRISP-DM,» IBM, [En ligne]. Available: https://www.ibm.com/docs/fr/spss-modeler/saas?topic=dm-crisp-help-overview. [Accès le 25 Mai 2024].

[47] IBM, «Qu'est-ce que l'algorithme des k plus proches voisins ?,» [En ligne]. Available: https://www.ibm.com/fr-fr/topics/knn. [Accès le 23 Mai 2024].

[48] IBM, «What is random forest?,» [En ligne]. Available: https://www.ibm.com/topics/random-forest. [Accès le 23 Mai 2024].

[49] IBM, «What is boosting?,» [En ligne]. Available: https://www.ibm.com/topics/boosting. [Accès le 23 Mai 2024].

[50] IBM, «What is a neural network?,» [En ligne]. Available: https://www.ibm.com/topics/neural-networks. [Accès le 23 Mai 2024].

[51] CNIL, «La CNIL, c'est quoi ?,» [En ligne]. Available: https://www.cnil.fr/fr/cnil-direct/question/la-cnil-cest-quoi. [Accès le 16 mai 2024].

[52] ANSSI, «Glossaire,» 21 Septembre 2024. [En ligne]. Available: https://cyber.gouv.fr/glossaire#:~:text=Ensemble%20des%20moyens%20mis%20en,et%20la%20lutte%20informatique%20offensive.. [Accès le 3 Novembre 2023].

[53] Crowdstrike, «EDR, MRD ou XDR,» 08 Août 2022. [En ligne]. Available: https://www.crowdstrike.fr/cybersecurity-101/endpoint-security/edr-vs-mdr-vs-xdr/. [Accès le 17 Mai 2024].

[54] MITRE, [En ligne]. Available: https://attack.mitre.org/. [Accès le 09 10 2023].

[55] Wikipédia, «Shellcode,» 2023 Décembre 2023. [En ligne]. Available: https://fr.wikipedia.org/wiki/Shellcode. [Accès le 22 Mai 2024].

[56] The OWASP Foundation, «OWASP Secure Coding Practices,» 2010.

[57] NIST, «Honeypot,» [En ligne]. Available: https://csrc.nist.gov/glossary/term/honeypot. [Accès le 22 Mai 2024].

[58] CERT-FR , «À propos du CERT-FR,» [En ligne]. Available: https://www.cert.ssi.gouv.fr/a-propos/. [Accès le 22 Mai 2024].

[59] scikit-learn, «scikit-learn,» [En ligne]. Available: https://scikit-learn.org/stable/. [Accès le 24 Mai 2024].

[60] Forum InCyber, «FORUM INCYBER,» [En ligne]. Available: https://europe.forum-incyber.com/le-forum/.

[61] CNIL, «Les missions de la CNIL,» [En ligne]. Available: https://www.cnil.fr/fr/la-cnil/les-missions-de-la-cnil. [Accès le 27 Mai 2024].

8. Liste des annexes

9. Annexes

9.1. Introduction et gestion de projet

Annexe 1- Répartition du temps

Répartition du temps.xlsx

9.2. Comptes-rendus de réunions

Annexe 2 - Note de clarification

Note de clarification suite au premier entretien.pdf

Annexe 3 - Compte-rendu septembre

Compte-rendu du mois de septembre.pdf

Annexe 4 - Compte-rendu octobre

Compte-rendu du mois d'octobre.pdf

Annexe 5 - Compte-rendu semaine 1

Compte-rendu semaine de projet 1.pdf

Annexe 6 - Compte-rendu janvier

Compte-rendu du mois de janvier.pdf

Annexe 7 - Compte-rendu février

Compte-rendu du mois de février.pdf

Annexe 8 - Compte-rendu semaine 2

Compte-rendu semaine de projet 2 (13-05 ; 17-05).pdf

Annexe 9 - Compte-rendu semaine 3

Compte-rendu semaine de projet 3 (20-05 ; 24-05).pdf

Annexe 10 - Compte-rendu semaine 4

Compte-rendu semaine de projet 4 (27-05 ; 31-05).pdf

9.3. Fiches de lectures

Annexe 11- Fiche de lecture 1

Fiche de lecture - CTI - Issue and Challenges.docx

Annexe 12- Fiche de lecture 2

Fiche de lecture - AI in CTI.docx

Annexe 13- Fiche de lecture 3

Fiche de lecture - Application Security through Sandbox .docx

Annexe 14- Fiche de lecture 4

Fiche de lecture - Artificial Intelligence to Improve Cybersecurity.docx

Annexe 15- Fiche de lecture 5

Fiche de lecture - Avast-CTU Public CAPE Dataset.docx

Annexe 16- Fiche de lecture 6

Fiche de lecture - Fashion Images Classification.docx

Annexe 17- Fiche de lecture 7

Fiche de lecture - SANS CTI 2021 survey.docx

Annexe 18- Fiche de lecture 8

Fiche de lecture - The unified kill chain.docx

Annexe 19- Fiche de lecture 9

Fiche de lecture - A Comprehensive Survey on Identification of Malware Types and Malware Classification Using Machine Learning Techniques .docx

Annexe 20- Fiche de lecture 10

Fiche de lecture - Algorithmes d'Intelligence Artificielle en Cybersécurité & Intégration en environnements contraints.docx

Annexe 21- Fiche de lecture 11

Fiche de lecture - Artifcial intelligence in cyber security research advances, challenges, and opportunities.docx

Annexe 22- Fiche de lecture 12

Fiche de lecture - CyTIME - Cyber Threat Intelligence ManagEment framework for automatically generating security rules.docx

Annexe 23- Fiche de lecture 13

Fiche de lecture - Machine learning for Security and the IoT.docx

Annexe 24- Fiche de lecture 14

Fiche de lecture - Performance Evaluation of Machine Learning Classifiers in Malware Detection.docx

9.4. Notes d'information

Annexe 25- Note d'information 1

Note d'informations - Cyber kill chain.docx

Annexe 26- Note d'information 2

Note d'informations - IoC.docx

Annexe 27- Note d'information 3

Note d'informations - SOC.docx

Annexe 28- Note d'information 4

Note d'informations - Open CTI.docx

Annexe 29- Note d'information 5

Note d'informations - STIX & TAXII .docx

9.5. Documentation

Annexe 30 - Documentation de l'installation de CAPEv2

Documentation CAPEv2.docx

10.Table des figures

11.Table des tableaux