Pr. Abdoulaye DEME

Dr. Djicknoum DIOUF

REPUBLIQUE DU SENEGAL

Un Peuple – Un But – Une Foi

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR DE LA RECHERCHE ET DE L'INNOVATION

UNIVERSITE GASTON BERGER DE SAINT-LOUIS

U.F.R DE SCIENCES APPLIQUEES ET DE TECHNOLOGIE

SECTION PHYSIQUE APPLIQUEE

Mémoire de fin d'études pour l'obtention du

Master 2 Ingénierie Électronique et Télécommunications (MIETEL)

Sujet : Amélioration de la détection/prévention d'intrusion (IDS/IPS) par des Techniques de Deep Learning

Présenté le 07/07/2022 par :

M. Jean Claude Joseph BADJI

Sous la direction de :

Dr. Chérif DIALLO

Jury :

Président :	M. Papa Alioune FALL
Examinateur :	M. Amadou Seidou MAIGA
Examinateur :	M. Moustapha SENE
Encadrant :	M. Chérif DIALLO

Année Académique 2020/2021

Résumé

À l'ère de la technologie, le paysage des cybermenaces s'élargit, les attaques deviennent plus furtives, rapides et intelligentes et la cybersécurité apparaît comme l'une des nouvelles préoccupations principales des organisations. Les techniques de détection traditionnelles deviennent plus inefficaces contre les menaces polymorphes et les attaques zero-day. La recherche s'est depuis orientée vers l'Intelligence Artificielle (IA) pour rattraper l'état de l'art et l'intelligence des attaques. Le Machine Learning (ML) a rapidement montré ses limites en raison du gros volume de données, de la haute dimensionnalité imposées par l'ère du Big Data et de la charge de travail sur l'extraction manuelle des caractéristiques. L'IDS basé sur ML a montré de mauvaises performances sur des données comme KDDCUP99 et ses améliorations qui ne sont pas si réalistes. Le Deep Learning (DL) et le jeu de données CSE-CIC-IDS2018 viennent à la rescousse pour résoudre de tels problèmes. Dans cette étude, nous proposons une classification du trafic par un CNN unidimensionnel et une détection d'anomalies par un auto-encodeur profond/empilé (DAE). Nous utilisons le CSE-CIC-IDS2018 pour évaluer les modèles proposés. Le F1-score est utilisé comme métrique principale pour la classification du trafic basée sur CNN car le rappel, la précision, le taux de fausses alarmes (FAR) et le taux de faux négatifs (FNR) sont importants dans le domaine de la cybersécurité. La métrique atteint une moyenne de 97,38 % (FAR et FNR très faibles) sans tenir compte de l'attaque par infiltration et de 87,33 % (FAR très élevé et FNR acceptable) lorsque le modèle est respectivement formé par type d'attaque et avec toutes les attaques combinées. En parallèle, le modèle DAE fonctionne bien contre presque toutes les attaques. Les deux modèles affichent des performances élevées et prometteuses.

Mots clés : Cybersécurité ; Système de détection d'intrusion ; Détection d'anomalies ; Deep Learning ; Réseau de neurones convolutionnels ; Auto-encodeur ;

Abstract

As cyberthreat landscape becomes larger, the attacks become stealthier, faster and smarter, cybersecurity emerges as one of the new main concerns for the organizations. Traditional techniques become more ineffective against polymorphic threats and zero-day attacks. Research has since been oriented towards AI to catch up with the attacks state of art and smartness. Machine Learning (ML) quickly shown its limits because of data amount and high dimensionality imposed by the big data era, and the workload on manual feature extraction. The ML-based IDS shown poor performances on data like KDDCUP99 and its improvements which aren't that realistic. Deep learning and CSE-CIC-IDS2018 dataset come at the rescue to solve all these problems. In this study, we propose a traffic classification by a one-dimensional CNN and an anomaly detection by a deep/stacked autoencoder (DAE). We use the CSE-CIC-IDS2018 to evaluate proposed models. F1-score is used as the main metric for CNN-based traffic classification because both recall, precision, False Alarm Rate (FAR) and False Negative Rate (FNR) are important in cybersecurity field. The metric reaches an average of 97.38 % (very low FAR and FNR) without considering infiltration attack and 87.33 % (very high FAR and acceptable FNR) when the model is respectively trained per attack type and with the whole dataset. In parallel, the DAE model performs well against almost all attacks. The two models show high and promising performance.

Key Words : Cybersecurity ; Intrusion Detection Systems ; Anomaly Detection ; Deep Learning ; Convolutional Neural Network ; Autoencoder ;

Remerciements

Je rends grâce à Allah de m'avoir donné la force d'achever ce travail.

Je remercie ma maman pour sa patience et son endurance. J'espère et je prie pour qu'elle reste à mes côtés jusqu'à la réussite de ses petits-enfants.

Je remercie ma famille, mes sœurs pour leur soutien et leur patience.

Je remercies tous mes amis, je leur souhaite une grande réussite dans la vie présente et future.

Je remercie les universités Cheikh Anta DIOP de Dakar et Gaston BERGER de Saint-Louis de m'avoir permis d'acquérir autant de connaissances.

Je m'excuse auprès de mes camarades de chambre au campus pour le dérangement occasionné. Je les remercie pour leur soutien et leur compréhension.

Je m'excuse auprès de mes camarades de classe pour tout désagrément causé. Je les remercie pour tous les moments de partage et je leur souhaite une grande réussite dans leur vie présente et future.

Je m'excuse auprès du corps professoral pour tout désagrément causé. Je les remercie pour toute la connaissance transmise et je leur souhaite une longue vie utile et une réussite dans leur vie future.

Je remercie mon encadrant, le cher professeur Chérif DIALLO de m'avoir donné l'opportunité de travailler sur un sujet aussi intéressant et tout le temps nécessaire pour achever ce travail.

Table des matières

Résumé **i**

Abstract **ii**

Remerciements **iii**

Liste des Figures **vii**

Liste des Tableaux **viii**

Liste des Abréviations **ix**

Introduction générale **1**

Chapitre 1 Vulnérabilités, Menaces et Risques **3**

1.1 Introduction 3
1.2 Objectifs de sécurité 4
1.3 Security Awareness 5
1.3.1 Origines des vulnérabilités 5
1.3.2 Menaces 6
1.3.3 Threats actors 7
1.3.4 Vecteurs d'attaques 8
1.3.5 Risques 9
1.4 Attaques et Intrusions 9
1.4.1 Malwares vs Exploits 9
1.4.2 Attaques 10
1.4.3 Processus d'intrusion 21
1.4.4 Menace Persistante Avancée 22
1.5 Conclusion 23

Chapitre 2 Contre-mesures de sécurité et Prévention **24**

2.1 Introduction 24
2.2 Analyse et management de risques 25
2.2.1 Analyse de risques 26

2.2.2 Évaluation et management de risques 29
2.3 Les contre-mesures de sécurité 33
2.3.1 Les contre-mesures administratives 34
2.3.2 Les contre-mesures techniques 35
2.4 Les firewalls de nouvelle génération ou Next Generation Firewall 35
2.4.1 Le chiffrement 36
2.4.2 Le contrôle d'accès - Access Control 38
2.4.3 Le filtrage de contenu 39
2.4.4 Les firewalls sans états et les firewalls avec états 40
2.4.5 Des firewalls de troisième génération et UTM vers les firewalls de nouvelle génération (NGFW) 42
2.4.6 Les Web Application Firewall 43
2.4.7 La sécurité des terminaux – Endpoint Security 44
2.4.8 La sécurité du cloud 46
2.5 Conclusion 46

Chapitre 3 Systèmes de détection/prévention d'intrusion 48
3.1 Introduction 48
3.2 Système de détection/prévention d'intrusion 49
3.2.1 Taxonomie des IDS/IPS 49
3.2.2 Stratégie de détection des IDS/IPS 55
3.2.3 Stratégie de déploiement des IDS/IPS 59
3.2.4 Stratégie d'évaluation des IDS/IPS 60
3.3 Challenge : la détection de l'inconnu 61
3.3.1 SIEM, Automatisation et TI 62
3.3.2 Honeypot/Honeynet et Sandboxing 63
3.3.3 Vulnérabilités zero-day et Bug Bounty 64
3.4 Limites des IDS et Stratégie de détection basée sur l'IA 65
3.4.1 Vulnérabilités et limites des IDS 65
3.4.2 Stratégie de détection basée sur l'IA 67
3.5 Conclusion 70

Chapitre 4 Stratégie de détection de l'inconnu par Deep Learning 71
4.1 Introduction 71
4.2 Étude de performance des différents modèles IDS basés sur le CNN et l'auto-encodeur pour la détection d'intrusion 72
4.3 Approches méthodologiques 74
4.3.1 EDA et pré-traitement 74
4.3.2 Modèles 75
4.4 Expériences 77

4.4.1 EDA 77
4.4.2 Pré-traitement des données 80
4.4.3 Résultats 82
4.5 Discussion 86
4.6 Conclusion 88

Conclusion générale **89**

Bibliographie **93**

Table des figures

1.1 Equifax data breach process 7
1.2 Vulnérabilités des serveurs web 15
1.3 Modèle du réseau GAN 21
1.4 Phases d'une menace persistance avancée 23

2.1 Les risques en fonction des éléments — vulnérabilités — menaces 27
2.2 Les contre-mesures de sécurité 34
2.3 Le modèle de sécurité d'accès au réseau 36
2.4 Les différentes zones d'un réseau 37
2.5 Public-Key vs. Symmetric-Key : Architecture 38
2.6 Le contrôle d'accès au réseau 39
2.7 Le content security dans la pile TCP/IP 40
2.8 Communication à travers un firewall sans état 41
2.9 Communication à travers un firewall avec état 41
2.10 Les produits de sécurité pour terminaux 46

3.1 La taxonomie des IDS 50
3.2 L'architecture des NIDS 51
3.3 Principe de détection des NIDS actuels 52
3.4 Emplacement des sondes NIDS 53
3.5 Schéma de fonctionnement de SNORT 55
3.6 La Stratégie de détection des IDS 70

4.1 Architecture du Deep Autoencoder 76
4.2 Architecture du 1D CNN 77
4.3 Distribution du trafic dans CSE-CIC-IDS2018 78
4.4 Corrélation des caractéristiques du CSE-CIC-IDS2018 80
4.5 Processus de l'expérience de la détection d'anomalies 81
4.6 Processus de l'expérience de la classification d'attaques 81
4.7 Courbe d'apprentissage du DAE 83
4.8 Matrices de confusion du CNN par type d'attaque 85
4.9 Matrice de confusion du CNN vs toutes les attaques combinées 87

Liste des tableaux

2.1 Les contrôles proposés par la norme ISO 27002 26
2.2 Risk Matrix/Risk Heat Map 31

3.1 Les règles de SNORT 56
3.2 Inconvénients des IDS 67
3.3 Limites des IDS 67

4.1 Information sur les caractéristiques CSE-CIC-IDS2018 78
4.2 Le pourcentage de trafic par type d'attaques 79
4.3 Pourcentages d'anomalies par seuil 83
4.4 Rapport de classification avec le CNN pour chaque type d'attaque 84
4.5 Rapport de classification avec le CNN contre CSE-CIC-IDS2018 88

Liste des Abréviations

AAA	Authentication - Authorization - Accounting
AES	Advanced Encryption Standard
ALE	Annualized Loss Expectancy
APT	Advanced Persistent Threat
ARO	Annualized Rate of Occurrence
AWS	Amazon Web Services
BYOD	Bring Your Own Device
CA	Certificate Authority
CIC	Canadian Institute for Cybersecurity
CNN	Convolutional Neural Network
CSE	Communications Security Establishment
DAE	Deep Auto-Encoder
DDoS	Distributed Denial of Service
DES	Data Encryption Standard
DL	Deep Learning
DMZ	Demilitarized Zone
DNN	Deep Neural Network
DNS	Domain Name System
DOM	Document Object Model
DTP	Dynamic Trunk Protocol
DoS	Denial of Service
EDR	EndPoint Detection Response
FAR	False Alarm Rate
FNR	False Negative Rate
FN	Faux Négatifs

FP Faux Positifs

FTP File Transfer Protocol

GAN Generative Advesarial Network

GRU Gated Recurrent Unit

HOIC High Orbit Ion Canon

IA Intelligence Artificielle

IDS Intrusion Detection System ou Système de Détection d'Intrusion

IMAP Internet Message Access Protocol

IPS Intrusion Prevention System ou Système de Prévention d'Intrusion

ISMS Information Security Management System

IoC Indicators of Compromission

IoT Internet of Things

LOIC Low Orbit Ion Canon

LSTM Long Shot Time Memory

MAE Mean Absolute Error

MD5 Message Digest 5

MITM Man In The Middle

ML Machine Learning

MSE Mean Squared Error

NGFW Next Generation Firewall

NIST National Institute Security Technology

NaN Not a Number

RADIUS Remote Authentication Dial-In User Service

RC4 Rivest Cipher 4

RNN Recurrent Neural Network

ROC Receiver Operating Characteristic

RSA Rivest – Shamir – Adleman

SIEM Security Information and Event Management

SMTP Simple Mail Transfer Protocol

SNAP-R Social Network Automated Phishing with Reconnaissance

SOAR Security Orchestration, Automation and Response

SPOF Single Point of Failure

SSH Secure Shell

TAP Test Access Port

TCP Transmission Control Protocol

TI Threat Intelligence

TLS Transport Layer Security

TN True Negative

TP True Positive

TTL Time To Live

UDP User Datagram Protocol

UTM Unified Threat Management

WAF Web Application Firewall

XSS Cross-Site Scripting

Introduction générale

De nos jours, les organisations sont exposées à des menaces, des intrusions et des attaques à cause d'internet et aux technologies. La cybersécurité devient alors une préoccupation majeure. Pour faire face à cette exposition, nous construisons essentiellement notre sécurité avec des politiques de sécurité définies en déployant des firewalls au périmètre et des Intrusion Detection System ou Système de Détection d'Intrusion (IDS) à l'intérieur du réseau. La détection d'intrusion est un ensemble de mécanismes et de pratiques utilisés pour détecter les erreurs pouvant entraîner une défaillance de la sécurité, et diagnostiquer les intrusions et les attaques. Un IDS est la mise en œuvre de la détection d'intrusion. L'IDS utilise des signatures d'attaques et des techniques basées sur les anomalies pour la détection [1]. L'IDS basé sur les signatures compare les données de trafic à une base de données de signatures d'attaques connues pour la détection. Les IDS plus récents utilisent la détection d'anomalies basée sur des techniques heuristiques en recherchant des anomalies ou des écarts par rapport à l'utilisation normale pour la détection.

En pratique, malgré leur utilité, les IDS souffrent de nombreux problèmes [2].

Problème 1
Le premier problème est le taux existant de Faux Négatifs (FN). Nous obtenons des FN lorsque l'IDS considère le trafic malveillant comme normal. Ils sont généralement dus à l'incapacité de l'IDS à détecter des attaques inconnues de type « zero-day ».

Les chercheurs en cybersécurité ont proposé une technique de détection d'anomalies pour réduire les FN, car il est difficile de maintenir constamment une base de données de signatures à jour. Cela donne à l'IDS la capacité de détecter des attaques inconnues en recherchant un écart par rapport à un profil d'utilisation normal.

Problème 2
Cependant, l'IDS basé sur les anomalies génère un taux élevé de Faux Positifs (FP). Les FP sont les alarmes générées par l'IDS lorsqu'il identifie le trafic normal comme malveillant. Les IDS basés sur les anomalies génèrent trop de FP, ce qui prend du temps et augmente la charge de travail des professionnels de la sécurité.

Problème 3
Un troisième problème non négligeable est l'incapacité des IDS actuels à gérer la quantité de données et le débit des réseaux actuels. Ajouté à ces problèmes, le paysage des cybermenaces ne cesse de s'étendre et les cyberattaques deviennent de plus en plus intelligentes et furtives.

Solution Pour remédier à ces inconvénients des IDS, la recherche s'est orientée sur les modèles et techniques d'IA via le Machine Learning (ML) pour améliorer le taux de détection des attaques connues et zero-day, mais aussi donner une optimisation sur la classification et le clustering des techniques d'attaques pour réduire davantage les fausses alarmes et les faux négatifs. Mais en raison du gros volume de données et de la grande dimensionnalité des données, certains modèles de ML comme la machine à vecteurs de support (SVM) ont montré leurs limites. Le Deep Learning (DL) reste le candidat idéal pour :

- Gérer la haute dimensionnalité et la quantité de données ;
- Extraire et apprendre automatiquement les caractéristiques du trafic réseau ;
- Et prédire en conséquence selon les labels ou le seuil.

Certaines études récentes [3, 4] ont montré que le DL pourrait être une solution fiable pour les IDS. La plupart d'entre eux ont proposé divers modèles DL intéressants mais ont utilisé des ensembles de données tels que KDD et NLS-KDD qui sont obsolètes. Le Communications Security Establishment (CSE) et l'Institut canadien pour la cybersécurité (Canadian Institute for Cybersecurity (CIC) dans le cadre d'un projet collaboratif fournissent l'ensemble de données le plus à jour. Le CSE-CIC-IDS2018 est un ensemble de données Big Data avec presque 17 millions d'échantillons et 15 classes de trafic (une classe de trafic normal et 14 attaques différentes). Il n'y a que peu d'études avec le CSE-CIC-IDS2018 et certaines d'entre elles ont montré de mauvaises performances sur la détection des anomalies ou sur la classification en raison du déséquilibre du nombre d'échantillons par classe dans l'ensemble de données.

Dans notre étude, nous proposons une détection d'anomalies non supervisée par un auto-encodeur profond ou Deep Auto-Encoder (DAE) et une classification multiclasse de trafic par un réseau de neurones convolutionnels ou Convolutional Neural Network (CNN). Nous étudions plus en profondeur l'ensemble de données CSE-CIC-IDS2018 par le biais d'une analyse exploratoire des données et nous utilisons une technique générale de pré-traitement pour nettoyer les données avant d'entraîner et de tester les modèles. Pour résoudre le problème de déséquilibre des données, nous utilisons une technique de suréchantillonnage aléatoire pour la classification multiclasse du trafic.

Chapitre 1

Vulnérabilités, Menaces et Risques

1.1 Introduction

Selon le National Institute Security Technology (NIST), la sécurité est la protection fournie à un système d'information dans le but d'atteindre les objectifs applicables de préservation de l'intégrité, de la disponibilité et de la confidentialité des actifs du système d'information (matériel, logiciel, firmware, information/données et télécommunications). Tout système comporte des failles, parfois inhérentes, inconnues, ou encore connues mais non corrigées, que l'on nomme vulnérabilités et qui ouvrent des portes vers le système aux nombreuses menaces existantes. Les attaques représentent l'exécution de menaces contre un individu ou une organisation. La surface d'attaque des organisations s'élargit de plus en plus avec une large variété d'attaques qui sont lançables partout où il y a une vulnérabilité. Il existe des vulnérabilités dans chaque couche du modèle TCP/IP dont les couches basses présentent peu de mécanismes de sécurité et encore plus sur les protocoles d'application qui ne sont pas normalisés. Les cybercriminels sont susceptibles d'exploiter toute vulnérabilité et parviennent même à réaliser des intrusions très sophistiquées comme les Advanced Persistent Threat. Il faut noter qu'en plus de toutes ces vulnérabilités, les membres d'une organisation sont les plus exposés contre les attaques comme le social engineering. À quoi bon perdre énormément de temps en essayant de contourner un firewall ou la défense d'une organisation si on peut utiliser la négligence et l'inconscience de ses membres en manière de sécurité pour pénétrer dans le réseau ?

L'existence de ces menaces combinées aux vulnérabilités mène souvent à des risques. Ces trois éléments forment un ensemble qui nous empêche constamment d'atteindre les objectifs de sécurité. Il est donc impérial de bâtir la sécurité des systèmes d'information en essayant d'atteindre ces objectifs. La sécurité de l'information fait alors référence aux différentes contre-mesures mises en place pour protéger les membres, les processus et la technologie de l'organisation, le réseau et les données qui y sont stockées ou qui y

transitent. La sécurité nous permet de protéger notre organisation des menaces extérieures provenant d'internet, mais aussi des menaces internes par la présence des insiders ou encore des menaces environnementales. Dans cette première partie, nous faisons une étude de ces éléments importants à comprendre qui nous empêchent en permanence d'atteindre les objectifs de sécurité. Cette étude nous permettra de poser les bases de la source de notre problématique en faisant dans la foulée une sensibilisation à la sécurité. Nous faisons une étude en profondeur d'une large variété d'attaques avant de conclure.

1.2 Objectifs de sécurité

Pour assurer la sécurité, la communauté dans le domaine de celle-ci a jugé nécessaire de définir des objectifs à atteindre par les systèmes. Les objectifs de la sécurité sont :

— La **confidentialité**
Avec cette propriété, les informations confidentielles ne sont pas disponibles ou divulguées à des personnes non autorisées. Donc tout individu non autorisé sera dans l'incapacité de lire l'information confidentielle.

— L'**intégrité**
L'intégrité est une exigence sur l'information qui ne peut pas être modifiée en une manière non autorisée et non détectée. En d'autres termes, si une partie de l'information a été modifiée par un utilisateur non autorisé ou d'une manière non autorisée, on peut détecter cette modification.

— La **disponibilité**
La disponibilité est une propriété essentielle avec laquelle un système informatique fonctionne normalement à tout moment, et que ses services ne soient pas refusés aux utilisateurs autorisés. Dans ce cas, lorsqu'un utilisateur autorisé demande l'accès à un service, le système doit répondre rapidement à cette requête.

Après cela, la communauté du domaine de la sécurité a senti le besoin d'ajouter d'autres concepts à ces objectifs afin de présenter une copie complète. Donc au-delà des trois objectifs fondamentaux cités plus haut, il y existe deux autres objectifs supplémentaires qui les accompagnent :

— L'**authenticité**
L'authenticité exige la propriété d'être authentique et de pouvoir être vérifiée et digne de confiance, la confiance dans la transmission, d'un message ou d'un expéditeur de message. Par exemple, l'authenticité d'une entité signifie que cette entité est ce qu'il clame être.

— L'**accountability**
Cet objectif concerne l'exigence de pouvoir tracer les actions d'une entité unique-

ment jusqu'à cette même entité et pas vers une autre. Les systèmes d'information doivent suivre et enregistrer les activités d'un utilisateur pour permettre plus tard des analyses forensiques dans le but de repérer les brèches de sécurité du système. L'accountability implique la non-répudiation, qui veut dire qu'un individu ne peut nier avoir fait quelque chose. Par exemple, si on envoie un message, il ne sera pas possible de nier cette opération d'envoi. Si l'accountability peut être achevé donc les actions de chaque entité peuvent être tracées uniquement vers cette entité, donc on est incapable de nier ce qu'on a eu à faire.

1.3 Security Awareness

La security awareness est l'un des éléments essentiels pour maintenir le cap sur les objectifs de sécurité. Les attaquants parviennent à contourner nos contre-mesures de sécurité en misant et en utilisant la négligence et l'insouciance des membres de notre organisation grâce à des techniques de phishing par exemple. Dans cette partie, nous allons traiter des éléments essentiels qu'il est impératif de connaître pour tous les membres d'une organisation pour la maintenir sécurisée.

1.3.1 Origines des vulnérabilités

Tout système, application, logiciel, réseau, chose conçue par l'homme (man made en anglais) de manière générale présente des faiblesses, des vulnérabilités. La norme ISO 27005 définit la vulnérabilité comme [5] :
Une faiblesse d'un élément ou d'un groupe d'éléments qui peut être exploitée par une ou plusieurs menaces, où un actif est tout ce qui a de la valeur pour l'organisation, ses opérations commerciales et leur continuité, y compris les actifs d'information qui soutiennent la mission de l'organisation. L'IETF RFC 4949 définit la vulnérabilité en tant que [6] :
Une faille ou une faiblesse dans la conception, la mise en œuvre ou le fonctionnement et la gestion d'un système qui pourraient être exploitées pour enfreindre la politique de sécurité du système.

Il existe de nombreuses façons pour les attaquants de trouver une manière de pénétrer un réseau. L'exploitation de vulnérabilités est courante parce que ces dernières laissent de grandes ouvertures facilitant l'accès à notre réseau ou système. Les attaquants peuvent profiter entre autres :

- D'autorisations laissées ouvertes permettant à n'importe qui d'accéder aux ressources d'une organisation. Par manque de configuration de sécurité, Verizon [7] a exposé accidentellement, en juin 2017, 6 millions d'enregistrements de données

dans un serveur de stockage Amazon S3 ;

- De compte administrateur ou root non sécurisé avec des identifiants par défaut ou des mots de passe vulnérables aux attaques par force brute ;
- Des bogues qui, souvent lorsqu'elles ont lieu, demandent de fournir des informations à l'équipe support. Il arrive de fournir trop d'informations qui peuvent être interceptées par les attaquants comme en décembre 2015 avec le site Web Patreon [8] ;
- De l'utilisation de chiffrement faible avec des clés de sécurité plus petite que 128 bits ou d'algorithme de chiffrement ou de hachage obsolète comme le Message Digest 5 (MD5) ou encore quelques-uns des suites de chiffrement du protocole Transport Layer Security (TLS) ;
- De l'utilisation de protocoles non sécurisés comme Telnet, FTP, IMAP, SMTP, etc. alors qu'il existe leurs versions sécurisées ;
- De paramètres comme les identifiants d'appareils ou d'équipements laissés par défaut. En effet, ces appareils pourront être contrôlés pour participer à des botnets pour lancer des attaques DDoS ;
- De ports ou de services réseau laissés ouverts ;
- D'une gestion incorrecte des mises à jour. En effet, un retard ou une absence de mises à jour peut causer d'énormes dégâts comme l'illustre le cas d'Equifax entre mai et juillet 2017 [9]. Le déroulement de l'attaque est illustré par la figure 1.1.
- De l'utilisation de plateformes et systèmes obsolètes comme les anciennes versions de Windows ou d'une quelconque application qui ne bénéficient plus de maintenance et des mises à jour.
- De vulnérabilités zero-day, car toute application, logiciel, système, réseau présentent des vulnérabilités inconnues. Les fournisseurs ont des équipes spécialement dédiées à la recherche de failles dans leurs produits, en plus des chercheurs en sécurité. Ces équipes sont en constante course pour devancer les attaquants à la découverte et à la correction des failles avant qu'elles ne soient exploitées [10].

1.3.2 Menaces

En sécurité, on appelle une menace tout élément susceptible de compromettre l'un des objectifs de sécurité. Il existe quatre catégories de menaces :

— Il y a l'**interception** d'information lorsque l'attaquant accède aux ressources ou se place entre les entités qui communiquent pour intercepter dans le but de copier et voler des informations. L'interception affecte la confidentialité ;

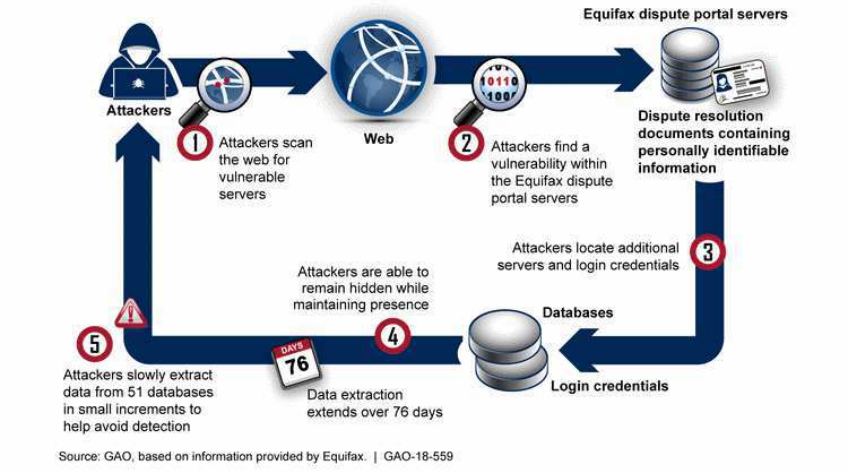


Figure. 1.1 : Equifax data breach process [9]

— Il y a la **modification** d'information lorsque l'attaquant accède non seulement aux ressources ou informations, mais les modifie sans autorisation à des fins de sabotage ou de fraude. La modification affecte l'intégrité ;

— Il y a l'**interruption** lorsque l'attaquant met en œuvre des mécanismes et réussit à couper une communication ou le fonctionnement d'un système, d'un service, etc. L'interruption affecte la disponibilité ;

— Il y a la **fabrication** d'information lorsque l'attaquant fabrique des informations frauduleuses et les insère dans une communication, dans une base de données, dans un réseau, dans un mail, etc.

1.3.3 Threats actors

On appelle un Threat actor, ou encore un Bad actor, toute entité qui se trouve derrière l'exécution ou l'occurrence de la menace et de la compromission de la sécurité d'un individu ou d'une organisation. Il existe un grand nombre d'acteurs qui ont une caractéristique en commun, mais des motivations différentes. Dans cette section, nous allons discuter des différentes catégories d'acteurs ainsi que certains de leurs attributs. Parmi les Threat actors, on retrouve :

- Le script kiddie qui compromet la sécurité juste pour amusement et n'est pas conscient des enjeux derrière ;
- Les insiders qui sont considérés comme les plus dangereux parce qu'ils sont internes aux organisations. La plupart des contre-mesures prises n'ont pas d'effet sur eux, car ils ont un accès direct et autorisé à plusieurs informations et ressources. Il peut s'agir d'employés, de collaborateurs, de fournisseurs (leurs équipements) commettant des erreurs par inadvertance ou de manière volontaire ;

- L'hactivist qui lance des attaques pour des causes sociopolitiques. Ses attaques sont souvent associées à un message social ou politique ;
- Les cyberwarriors qui sont appelés ainsi parce qu'ils piratent pour leur gouvernement ;
- Les groupes de cybercriminels qui forment une organisation et ont pour principal but le gain financier ;
- Et enfin les concurrents qui peuvent embaucher des cybercriminels pour espionner ou voler des informations.

1.3.4 Vecteurs d'attaques

Les vecteurs d'attaque sont les méthodes utilisées par l'attaquant pour accéder à notre système. Les attaquants passent beaucoup de temps en phase de reconnaissance pour recenser les vulnérabilités. Souvent, c'est la nature des vulnérabilités trouvées qui détermine les vecteurs d'attaque utilisés lors d'une attaque ou encore la qualité de la sécurité de la cible. D'autres misent sur la négligence et l'insouciance de la cible ou des membres de l'organisation cible. Dans cette section, nous allons discuter des différents vecteurs utilisés par les attaques. Parmi les possibilités qu'il peut avoir, un attaquant peut :

- Accéder directement aux équipements pour brancher un périphérique de stockage externe pour collecter des données ou alors installer un enregistreur de frappe (keylogger) et récupérer après les informations saisies. Il peut aussi causer l'interruption de ces équipements. Ce vecteur contourne les technologies de cybersécurité existantes ;
- Utiliser le niveau de sécurité faible des réseaux sans fil pour insérer un point d'accès (AP) clandestin qui peut être une porte pour les cybercriminels ou un evil twin. Un evil twin est utilisé pour usurper les fonctions des AP légitimes. Il peut collecter les données des utilisateurs pour l'attaquant qui aura la possibilité de voler les données, d'usurper des identités, etc. ;
- Utiliser le courriel qui est à ce jour le vecteur d'attaque le plus efficace. Ce vecteur d'attaque est souvent associé à des attaques comme le phishing, le social engineering ou encore la jointure de logiciels malveillants tels que les virus, les worms, etc. ;
- Utiliser la chaîne d'approvisionnement pour attaquer un individu ou une organisation. En effet, cette chaîne peut présenter des vulnérabilités et le fournisseur même peut exploiter ces vulnérabilités ;
- Utiliser les informations partagées sur les réseaux sociaux pour deviner les mots de passe, réinitialiser ou encore contourner l'authentification multifacteur.

1.3.5 Risques

Selon plusieurs rapports publiés, les vulnérabilités et les attaques ont de très grands impacts sur les individus ou sur les organisations. Il faut noter que ces impacts sont le plus souvent financiers comme dans le cas du réseau SWIFT [11]. En effet, il existe également d'autres risques comme :

- La perte de données ;
- Le vol d'identité comme dans le cas d'Equifax [12] ;
- L'impact sur la réputation comme dans le cas de Uber [13] ;
- La perte de disponibilité.

Ces coûts et ces effets des vulnérabilités nous poussent en matière de sécurité à nous assurer que nos systèmes sont aussi sûrs que possible.

1.4 Attaques et Intrusions

1.4.1 Malwares vs Exploits

La plupart du temps, les attaquants utilisent des virus, des Worms, des Trojan horses, des ransomwares, des adwares, des spywares, des spams, des bots, etc. pour attaquer nos systèmes. Ces malwares ne dépendent pas de notre système, mais d'un manque de security awareness ou de la négligence. Ces malwares sont des exécutables qui ne dépendent que d'un clic de l'utilisateur, ou juste d'être dans le système pour se lancer et l'infecter. En réalité, un malware peut avoir une famille composée de millions de variants qui font généralement la même chose une fois dans notre système. Ce type de malware est appelé polymorphique. Un malware est l'outil le plus simple utilisé par un hacker. Cependant, les utilisateurs de nos jours sont de plus en plus conscients de l'importance de la sécurité. En plus des contre-mesures déployées, plusieurs organisations forment les membres sur les enjeux de la sécurité. Il est donc plus difficile de piéger un utilisateur avec des attaques de type social engineering. Avec cela, les hackers sont contraints d'étudier les vulnérabilités des systèmes afin de les exploiter. Ils développent des scripts ou applications qui tirent parti d'une vulnérabilité pour prendre le contrôle d'un système, modifier des données ou désactiver un service [14]. Ce genre de script est appelé *exploit*[1]. Un *exploit* est dépendant de la vulnérabilité et il peut regrouper plusieurs techniques d'exploitation. Les techniques d'exploitation au nombre de 20 à 30 environ sont les actions effectuées par l'*exploit*. Un *exploit* a besoin de 2 à 6 techniques en moyenne pour être efficace [15].

1. exploit : mot anglais utilisé pour designer un script utilisé pour exploiter une vulnérabilité.

1.4.2 Attaques

Une attaque représente l'exécution d'une ou de plusieurs menaces ou encore l'exploitation d'une faille. Une attaque peut être un ensemble de techniques mises en œuvre pour exécuter plusieurs menaces ou pour exploiter plusieurs vulnérabilités à la fois. Il existe une très grande variété d'attaques et dans cette section, nous étudions quelques classes d'attaques en examinant les attaques les plus couramment utilisées selon [16]. Nous donnons une grande importance aux attaques web parce qu'elles constituent le réel problème, et que les firewalls ont beaucoup plus de mal à bloquer ces types d'attaques.

1.4.2.1 Social Engineering

Le social engineering est l'art de la tromperie. Dans cette famille d'attaques, l'attaquant utilise la tromperie pour essayer de gagner la confiance des cibles dans le but de leur soutirer des informations sensibles, d'installer des malwares ou de compromettre la sécurité de celles-ci. Le social engineer apprend à connaitre ses cibles en collectant des informations, décide de la technique appropriée à utiliser, utiliser les ruses et la naïveté humaine pour gagner la confiance des cibles, les manipule dans le but de divulguer des informations sensibles ou de violer les politiques de sécurité définies par les organisation des cibles. Les techniques de social engineering s'attaquent directement aux utilisateurs plutôt qu'à l'infrastructure de l'organisation. C'est pourquoi elles sont très difficiles à gérer. Les techniques du social engineering sont :

- Le Baiting ;
- Le Scareware ;
- Le Pretexting ;
- Le Phishing, le Whale - phishing, le Vishing ;
- Le Spear Phishing qui est un cas particulier du phishing et cible un individu ou une organisation spécifique ;
- Le Water Holing, etc.

1.4.2.2 Attaques basées sur les vulnérabilités des protocoles

La plupart des appareils connectés communiquent selon le modèle TCP/IP. Il y a des protocoles à chaque couche de ce modèle et la plupart de ces protocoles présentent des vulnérabilités qui sont exploitées par des cybercriminels.

> **Attaques sur la couche liaison de données**
Dans cette couche, nous retrouvons les protocoles Ethernet, ARP, les étiquettes VLAN.

- Les entêtes Ethernet ou encore MAC ne sont protégés par aucun mécanisme de sécurité. Cela donne à un cybercriminel la possibilité de forger ces entêtes et d'usurper un hôte du réseau (MAC Spoofing). Avec cette vulnérabilité, il peut intercepter ou interrompre tout le trafic destiné à l'hôte usurpée compromettant ainsi la confidentialité ou la disponibilité de la communication ;
- L'attaquant peut aussi forger plusieurs différentes entêtes et submerger le switch jusqu'à remplir sa table MAC (MAC Flooding). Dès que la table MAC du switch est remplie, il change de comportement et commence à diffuser tout le trafic reçu sur tous ses ports comme un hub. Avec ces vulnérabilités, l'attaquant peut capturer tout le trafic tout en empêchant au switch de remplir son rôle de commutation compromettant ainsi la confidentialité et la disponibilité de la communication ;
- Le protocole ARP est utilisé sur la couche 2 pour créer un mappage entre l'adresse IP et l'adresse MAC des hôtes. Étant donné que les premières requêtes ARP sont diffusées sur le réseau, un attaquant peut répondre en premier à celles-ci, ou plus facile pour lui s'autoproclamer comme un autre hôte grâce au mécanisme de gratuitous ARP. L'exploitation avec succès de ces deux vulnérabilités (ARP poisoning) peut avoir comme conséquence l'interception du trafic vers l'hôte usurpé ou encore le déni de service ;
- Les entêtes VLAN sont utilisés pour isoler des hôtes dans différents LAN virtuels par ports, par adresse MAC, par protocole de niveau 3, par réseaux IP. Avec le Dynamic Trunk Protocol (DTP), il est possible de définir automatiquement les liens Trunk. Avec ce protocole, les switches annoncent sur les liens leur capacité d'être trunk. Avec le mode dynamic desirable de DTP, un attaquant peut se déclarer comme un switch et négocie avec les autres switches pour partager un lien trunk. Si ces switches ont le protocole DTP actif, alors l'attaquant aura la possibilité d'envoyer toutes ses trames à l'ensemble des vlans ou d'intercepter les communications de ces différents vlans. Les attaques MITM et les dénis de services sont possibles dans ces cas ;
- Étant donné que l'assignation des vlans est manuelle, une erreur de configuration peut causer l'assignation d'un ID du vlan natif des trunks à d'autres vlans où il y a des hôtes. Un attaquant qui se trouve dans un VLAN configuré comme natif par erreur peut attaquer un autre VLAN en utilisant une double encapsulation de tag et en rajoutant le tag de la victime.
- Etc.

> **Attaques sur la couche réseau**

La couche réseau identifie chaque hôte du réseau et assure la connectivité entre différents réseaux locaux. Cette couche n'a comme fiabilité que le best effort et les entêtes checksums qui sont utilisés pour détecter les erreurs.

- Étant en mode connectionless, l'attaquant peut usurper l'identité d'un hôte en forgeant l'entête IP (IP spoofing). Avec cette usurpation, le cybercriminel ne recevra pas les réponses, mais peut rester pendant longtemps dans le réseau avec la possibilité d'inclure une charge malveillante dans ses paquets. Il peut aussi déclencher un déni de service par réflexion sur l'hôte usurpé qui va recevoir toutes les réponses qui peuvent être amplifiées par l'utilisation des protocoles de la couche applicative aux requêtes lancées par l'attaquant.
- Le protocole ICMP qui est conçu à la base pour diagnostiquer les réseaux en permettant de récupérer les informations opérationnelles sur les statuts des réseaux, peut être utilisé par un attaquant pour des taches de reconnaissance avec le ping sweep (multiples ping pour scanner), le traceroute (découverte de la topologie) ou encore le fingerprinting basé sur le time Time To Live (TTL) (décompte du nombre de routeurs et identification du type d'équipement). Ce protocole peut aussi être utilisé pour lancer une attaque DoS nommée smurf en usurpant une adresse IP et en envoyant plusieurs pings à un hôte ou en diffusion.
- Un attaquant peut aussi utiliser un ICMP redirect pour rediriger les paquets vers lui. Si cette opération réussit, il peut alors bloquer les paquets causant un déni de service ou regarder le contenu avant de transférer vers la destination (attaque MITM).
- Etc.

> **Attaques sur la couche transport**

Les protocoles de transport permettent la communication entre processus qui sont identifiés par des numéros de port renseignés dans l'entête de la couche transport. Il existe deux protocoles de transport que sont User Datagram Protocol (UDP) (non orienté connexion) et Transmission Control Protocol (TCP) (orienté connexion et prenant en charge plusieurs mécanismes de fiabilité). Ces protocoles omniprésents présentent plusieurs vulnérabilités que les attaquants peuvent exploiter.

- Du fait du manque de mécanisme d'authentification sur les segments TCP lors des échanges dans une session, un attaquant peut usurper l'adresse IP de l'un des hôtes dans les échanges et deviner/prédire le numéro de séquence dans les échanges pour détourner la session. Cela peut avoir comme impact l'interruption de la session avec un flag RESET (RST) ou l'inclusion de charges malveillantes dans les échanges de segments.
- Il peut aussi causer un déni de service en exploitant la vulnérabilité du dépassement de tampon (Buffer overflow) en envoyant constamment des paquets SYN dans le but d'ouvrir de manière continue de nouvelles connexions jusqu'à épuisement de ressources. Les requêtes des utilisateurs normaux seront alors bloquées. Il s'agit d'attaques DoS/DDoS par épuisement de ressources.

- Dans le cas d'UDP, l'absence de connexion et l'absence de contrôle de congestion peuvent être exploitées pour envoyer un grand nombre de segments afin d'épuiser la bande passante du lien vers la victime. Il s'agit d'attaques DoS/DDoS par épuisement de bande passante.
- L'attaquant peut exploiter la présence de réflecteurs comme les résolveurs pour submerger un hôte dont il a usurpé l'adresse. Cette attaque, où les réponses aux requêtes de l'attaquant sont envoyées à la victime, est appelée backscatter. En utilisant les vulnérabilités d'UDP, l'attaquant peut, à travers les réflecteurs, submerger la victime avec le volume des réponses largement plus grand que celui des requêtes. C'est une attaque DoS/DDoS par amplification UDP souvent utilisée avec le protocole DNS. L'impact est proportionnel au nombre de résolveurs utilisés.
- Etc.

> **Attaques sur la couche application**
>
> La couche application présente une très grande diversité de protocoles, donc une très grande variété de vulnérabilités. Contrairement aux autres couches, il n'y a pas de management d'application à l'interface ni un modèle de traitement générique. Ce qui a augmenté la surface d'attaque déjà existante. Nous allons examiner quelques attaques basées sur les applications courantes :
>
> - Le protocole Domain Name System (DNS) est utilisé pour résoudre des noms de domaines en adresse IP étant donné que les noms de domaines sont plus faciles à retenir et qu'un service peut être hébergé sur plusieurs serveurs. Les serveurs DNS utilisent une mémoire cache pour sauvegarder les entrées des domaines où ils n'ont pas d'autorité. Cette mémoire cache peut-être «empoisonner» par un attaquant avec de fausses associations IP — Nom de domaine en devinant le code (TXID) utilisé pour identifier les requêtes DNS. Il peut en résulter une usurpation de l'identité d'un hôte, d'un site web, ou d'une attaque MITM. Grâce à cette attaque, un cybercriminel peut collecter des informations sensibles de l'utilisateur qu'il a redirigé dans son site web. Il peut aussi détruire la réputation d'un tiers en commettant des crimes en son nom ou en détruire son image en substituant le site à un autre.
> - Plus de détails sur certaines attaques de cette couche seront donnés dans la section des attaques web.

1.4.2.3 Attaques par force brute

Les attaques par force brute ont pour méthode la simple combinaison d'identifiants et de mots de passe de compte utilisateur à partir de chaînes de caractères simples, de dictionnaire, de mots de passe antérieurs pour essayer de deviner les bons identifiants et

gagner un accès sur les comptes utilisateurs. On peut retrouver une attaque par force brute contre tout protocole ou tout réseau ou encore tout site web qui utilise des identifiants pour accéder aux comptes ou aux ressources. Les attaques par force brute de nos jours sont plus redoutables et efficaces, car ils sont exécutés par des bots. Parmi les attaques brutes force, on peut citer des attaques comme le Secure Shell (SSH) brute force, le File Transfer Protocol (FTP) brute force, le brute force web, brute force XSS, etc. Les attaques par force brute peuvent aussi cibler les clés de sécurité de chiffrement.

1.4.2.4 Botnet

Un botnet malveillant est un réseau de serveurs, d'ordinateurs, d'équipements réseau, d'appareils IoT sous le contrôle d'un attaquant à cause d'infection par malware. Les attaquants peuvent utiliser le modèle client-serveur, avec des topologies étoilées ou hiérarchiques, ou le modèle Peer to Peer dans les botnets. Dans le modèle client-serveur, ils installent des agents sur les bots et les contrôlent avec des canaux cachés, command & control (C & C). Il existe plusieurs types d'attaques de type botnets :

- Les botmasters où on a un nœud central qui établit des connexions C & C avec ses sujets et les contrôler à distance. Le botmaster peut alors utiliser les ordinateurs sous son contrôle pour lancer des attaques coordonnées telles que les DDoS ;
- Les réseaux de zombies qui utilisent les machines connectées à internet pour effectuer des tâches malveillantes ;
- Les réseaux spambots utilisés pour distribuer des spams, de la publicité aux cibles ;
- Etc.

1.4.2.5 Attaques par déni de service (DoS, DDoS, ransomware)

Le Denial of Service (DoS) est un cybercrime utilisé par les attaquants pour rendre un service inaccessible pour les utilisateurs normaux. Ce type d'attaque peut être utilisé par un cybercriminel pour le gain, par un hacktivist pour faire passer un message ou pour exprimer son désaccord contre un tiers, par un concurrent pour récupérer un marché, etc. Il est possible de lancer des attaques DoS sur toutes les couches de TCP/IP. Il existe généralement deux manières de lancer un déni de service :

- Un DoS par l'épuisement de la bande passante en envoyant un grand volume de trafic qui congestionne le lien vers la cible. C'est une attaque basée sur le volume et l'attaque par amplification DNS en est un bon exemple ;
- Un DoS par l'épuisement des ressources en envoyant un grand nombre de requêtes à la victime. Ce type d'attaque est basée sur le protocole TCP et l'attaque par SYN flood en est un exemple ;

- Dans une attaque DoS, le cybercriminel utilise une seule machine pour rendre un service inaccessible aux clients alors que dans le cas d'une attaque Distributed Denial of Service (DDoS) il peut utiliser des milliers voire des millions de machines grâce aux botnets ou à des coordinations.
- Dans une attaque ransomware, l'attaque chiffre les données de la victime et demande une rançon affichée à l'écran pour fournir la clé de déchiffrement. Cette attaque peut toucher les périphériques externes branchées sur les appareils cibles lors de l'attaque.

Il faut noter que dans les attaques DoS/DDoS par épuisement de ressources, il y a celles de la couche application qui sont plus difficiles à détecter et qui submergent les serveurs. Des outils, comme Low Orbit Ion Canon (LOIC) et High Orbit Ion Canon (HOIC), Slowloris sont souvent utilisés pour lancer ces dernières attaques DoS/DDoS.

1.4.2.6 Attaques Web

Avec internet, l'utilisation des applications web a explosé avec une grande diversité qui n'est soumise à aucune règlementation au niveau mondial contrairement aux autres couches du modèle TCP/IP. De ce fait, ces applications présentent plusieurs vulnérabilités sur chaque nœud du schéma serveur/client et subissent un très grand nombre d'attaques (figure 1.2). Cela s'est multiplié avec l'arrivée des business websites (e-commerce, e-Banking, etc.). Dans cette section, nous allons examiner les quelques attaques web les plus fréquentes.

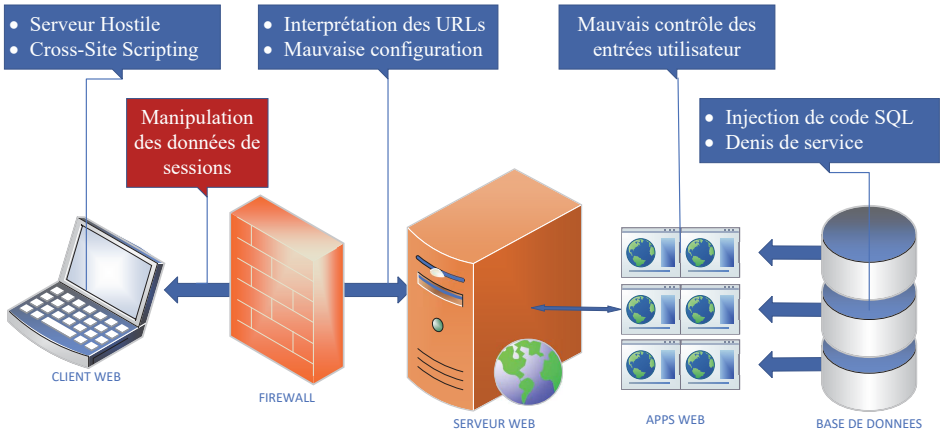


Figure. 1.2 : Vulnérabilités des serveurs web

— **L'attaque Cross-Site Scripting (XSS)**
Lors de cette attaque, l'attaquant écrit et transmet des scripts via des liens aux cibles dans le but de compromettre les comptes utilisateurs, de voler des données ou encore d'installer un malware. Cette technique utilise les langages de Scripting, notamment JavaScript qui peut accéder aux cookies, aux tokens de sessions, à tout

objet dont la page web a accès, mais aussi aux données de localisation, d'utilisateurs, au microphone, aux webcams grâce aux interfaces de programmation d'applications (API) de HTML 5, aux objets Document Object Model (DOM). L'attaquant peut utiliser JavaScript avec du contenu malveillant pour lancer ses attaques. Il existe plusieurs types de XSS :

- Le type reflected XSS dans lequel l'attaquant utilise ses liens et mails malveillants. Il incite les utilisateurs à lancer des requêtes à partir de ses liens qui contiennent les scripts malveillants. Dans ce cas, les clients utilisant ces liens envoient des requêtes au serveur web qui renvoie une réponse HTTP incluant cette même charge utile de l'attaquant. Cette charge utile est alors exécutée dans le navigateur du client ;
- Le type stored XSS dans lequel l'attaquant trouve le moyen de stocker sa charge utile dans la base de données de l'application. Ce dernier sert la charge utile de l'attaquant dans le navigateur de l'utilisateur qui exécute les scripts malveillants lors de sa navigation sur la page infectée.
- Le type XSS basée sur DOM dans lequel l'attaquant profite de la manipulation des objets DOM par l'application web pour injecter sa charge utile. La charge utile s'exécute seulement lorsque les données écrites dans l'objet DOM sont lues par le navigateur client.

Lorsque ces *exploits* JavaScript sont combinés aux techniques de social engineering, elles peuvent aboutir au vol de cookies, au keylogging, au vol d'identité, à l'installation de cheval de Troie, etc.

— **L'attaque Path Traversal**

Les attaques Path Traversal ciblent les applications web notamment le dossier racine web, dans le but d'obtenir un accès indésirable aux fichiers ou répertoires du serveur ciblé. L'attaquant en injectant des modèles de mouvement dans le répertoire du serveur peut être en mesure de remonter dans la hiérarchie. En cas de succès, l'attaquant peut supprimer des fichiers et compromettre l'accès au site, modifier et également compromettre les fichiers de configuration, les bases de données et d'autres sites web ainsi que les données sur le même serveur..

— **Les attaques Man In The Middle (MITM)**

Dans les attaques MITM, l'attaquant exécute la menace de type interception dans un réseau ou dans ce cas-ci entre le serveur et le client. Il écoute clandestinement et collecte toutes les informations échangées. Mais ce type d'attaque ne se limite pas aux écoutes clandestines, car dans certains cas, l'attaquant peut modifier le contenu, causer une interruption de la communication ou encore voler les cookies d'une session. Les attaques MITM sont très efficaces lors des échanges notamment de clés de session dans le processus de chiffrement.

— **Les attaques par injection de codes**
Ce sont des attaques qui sont assez courantes, notamment le SQL injection qui s'attaque à la base de données d'un site web. Dans une injection SQL, l'attaquant peut par exemple inclure une opération logique qui est toujours vraie ou utiliser des caractères spéciaux qui trompent le serveur ou l'authentification. Ces attaques sont les plus dévastatrices pour les sites web. L'injection de codes peut permettre de gagner un accès aux bases de données, d'y effectuer des modifications ou encore de compromettre l'application.

— **Les attaques Phishing**
Le phishing fait partie de la catégorie des attaques de type social engineering. Cependant, les attaquants utilisent tout le temps les sites web pour lancer ce type d'attaque. Ils peuvent utiliser des liens vers des sites frauduleux, les emails, etc.

— **Les attaques Fuzzy et Zero-day**
Le fuzzing est une technique utilisée dans les tests d'intrusion pour trouver des failles dans une application web, dans un système d'exploitation ou dans un réseau. Cette technique peut également être utilisée de la même manière par des attaquants pour trouver et exploiter des vulnérabilités dans des applications web par exemple. Le principe est d'injecter beaucoup de données dans l'application web pour la faire planter pour ensuite identifier les points vulnérables avec un outil Fuzzy. Les attaques zero-day étant des extensions des techniques de fuzzing, représentent les nouvelles de types de menaces. Nous pouvons avoir un cas où l'attaquant peut obtenir des informations sur les correctifs de sécurité à venir. Dans ce cas, il exploite ces failles spécifiques avant la publication des correctifs. L'autre possibilité est un attaquant qui a obtenu des informations sur la version du correctif et cible donc tout système obsolète.

1.4.2.7 Attaques basées sur l'IA

L'utilisation malveillante de l'IA augmente la vitesse d'exécution, le taux de réussite et augmente les capacités des cyberattaques [17]. Cette utilisation a pour effet :

- L'automatisation des processus d'attaque sans l'intervention humaine ;
- La facilitation des attaques qui sont devenues plus efficaces et efficientes ;

Il est donc important pour nous d'être à jour sur les nouvelles tendances dans la cybercriminalité pour mener des actions défensives appropriées. Des évènements comme le Cyber Grand Challenge de DARPA en août 2016 [18] ont proposé des illustrations de l'automatisation de la génération d'*exploits* et l'automatisation de processus d'attaque. Dans un futur proche avec l'utilisation de l'IA, nous assisterons probablement à :

- L'expansion des menaces existantes à l'origine des cyberattaques jusque-là très difficiles à mettre en œuvre ;
- L'introduction de nouvelles menaces à l'origine des cyberattaques irréalisables par l'homme ;
- Le changement du caractère typique des menaces qui impliquent de nouveaux attributs d'attaques finement ciblées, automatisées, très efficaces, à grande échelle et avec une quasi-impossibilité de traquer la source.

> **Étude de cas 1 : Next-Generation malware**
> Les attaquants peuvent utiliser des techniques de deep learning pour rendre ce nouveau genre de malwares furtif et capable de reconnaître des cibles spécifiques qui vont déclencher l'attaque.

Exemple 1 : DeepLocker

Le DeepLocker est un malware qui utilise des techniques de dissimulation et est destiné à des cibles spécifiques. Le DeepLocker est conçu avec deux réseaux de neurones profonds (DNN). Le premier DNN sert à dissimuler le malware en cryptant sa charge utile. Ainsi, les systèmes de sécurité comme les antivirus ne pourront pas le détecter. Chez les victimes, le DeepLocker commence à collecter des informations comme l'identification, la géolocalisation, la reconnaissance vocale et faciale. Ces informations vont servir de données d'entrée pour le deuxième DNN qui sert à générer la clé pour le décryptage de la charge utile. Lorsque la victime lance l'application, celle-ci alimente par exemple par la caméra le deuxième DNN. Si cette victime est la bonne cible, la charge utile du malware est décryptée et s'exécute sinon elle reste inactive. Ce genre de malware peut être utilisé pour contrôler un système à distance une fois que la bonne cible est identifiée. À ce jour, il n'y a pas de défense spécifique mise en œuvre pour le DeepLocker, mais restreindre l'accès aux capteurs de nos systèmes pour les applications pourrait nous aider.

Exemple 2 : Smart Malware

Les attaques traditionnelles provoquent des incohérences d'altération qui entraînent une forte probabilité de leur détection contrairement à un nouveau type de malwares appelé smart malwares. Ce nouveau type de malware a la capacité d'apprendre des stratégies d'attaque à partir des données de mesures d'un dispositif et de se propager furtivement dans le système cible pour y provoquer des pannes. Le malware accède à la base des données de mesures du dispositif. Il peut en déduire automatiquement des stratégies d'attaque afin de choisir un moment opportun pour déclencher l'attaque. L'élaboration des stratégies d'attaque passe par la détermination des caractéristiques de chaque mode de fonctionnement du dispositif à attaquer à l'aide d'une méthode de

classification de Machine Learning (k-means par exemple). Ensuite, ce malware intelligent analyse les paramètres en étudiant leur corrélation pour une bonne sélection de paramètres susceptibles d'entraîner des défaillances. Ce genre de malware est souvent utilisé pour attaquer des systèmes de contrôles de l'environnement [17]. Nous pouvons minimiser les effets de ce malware en déployant un système de détection d'intrusion avec des politiques de sécurité stricte, en utilisant une authentification multifacteur.

> **Étude de cas 2 : Synthèse de voix**
La synthèse de voix par IA est à l'origine de nouveaux types de fraudes comme l'imitation de la voix d'un client pour collecter ses données de transactions bancaires. Ce nouveau type d'attaque basée sur la synthèse vocale peut aussi constituer une fraude contre les processus de sécurité biométriques.

Exemple : Stealthy Spyware

Traditionnellement, un attaquant utilisant les attaques basées sur la synthèse vocale lançait leur attaque par haut-parleur la nuit durant le sommeil des cibles pour ne pas attirer l'attention. Avec l'IA, des attaques comme le stealthy spyware sont capables de synthétiser les mots clés d'activation de l'attaque de manière furtive pour ensuite décider du moment critique afin d'envoyer des commandes vocales malveillantes à l'assistant vocal de notre smartphone pour lui ordonner d'exécuter des tâches sans être détecté. Le stealthy spyware utilise un outil-espion, pour contrôler le microphone, qui réussit à avoir les autorisations suffisantes au moment de son installation. L'attaque utilise les quatre modules que contient l'outil stealthy spyware :

- Le module de surveillance de l'état et enregistrement vocal (SMVR) sert à contrôler l'état du microphone et a démarré l'enregistrement dès que le microphone est activé ;
- Le module Activation Voice manipulation (AVM) qui traite les enregistrements pour synthétiser des mots clés d'activation en intégrant des techniques de Natural Language Processing (NLP) ;
- Le module de détection intelligent d'environnement (IED) qui détermine le moment et le volume sonore optimaux en utilisant les capteurs du smartphone pour lancer l'attaque ;
- Le module Voice Assistant Activation and Control (VAC) pour lancer la voix d'activation avec un certain volume en fonction des résultats de l'IED.

En résumé, l'attaquant peut synthétiser des commandes vocales pour attaquer tout système de sécurité basée sur la parole et les assistants vocaux en imitant la voix des utilisateurs cibles et en s'adaptant à l'environnement.

> **Étude de cas 3 : Attaques basées sur les mots de passe**

Exemple 1 : Attaques par force brute de nouvelle génération

Parmi les attaques basées sur les mots de passe, nous avons une nouvelle génération d'attaque par force brute basée sur l'IA. Ce genre d'attaques a pour caractéristique la modification de la construction du dictionnaire d'attaque en utilisant des processus de Machine Learning. Les next-generation force brute attacks comme on les appelle génère de nouveaux mots de passe basés sur ceux qui sont antérieurs en utilisant un algorithme de Machine Learning comme les réseaux de neurones récurrents (RNN). Khoa Trieu et Yi Yang [13] ont entraîné un RNN open source (Torch-RNN) par des séquences de mots de passe captures antérieurement. Le RNN utilise la reconnaissance de patterns [2] pour mettre à jour son hidden state [3] et ainsi modéliser la distribution du caractère suivant dans la séquence en fonction des caractères précédents. Ce processus est utilisé pour prédire le caractère suivant, pour générer de nouveaux mots de passe et pour construire un dictionnaire intelligent. Ce type d'attaque est destiné à craquer les mécanismes d'authentification avec identifiant et mot de passe. La meilleure manière de se défendre contre ce genre d'attaques est d'utiliser une authentification multifacteur et des mots de passe complexe.

Exemple 2 : PassGAN

PassGAN est une nouvelle approche pour générer des suppositions de mots de passe de grande qualité sans aucune intervention humaine. Il a comme caractéristiques la capacité d'apprentissage des fuites de mots de passe précédents et d'extraction des propriétés et structures de façon autonome pour identifier les mots de passe hautement probables. Un réseau Generative Advesarial Network (GAN) (figure 1.3) peut être entraîné par un ensemble de données contenant des échantillons de mot de passe réels pour générer des mots de passe automatiquement. La générative G du GAN est entraînée avec un vecteur de bruit z et génère une séquence de vecteurs G(z). Cette séquence de faux mots de passe et l'ensemble des mots de passe réels constituent les entrées de la discriminative D qui doit apprendre pour distinguer les faux mots de passe des vrais.

> **Étude de cas 4 : Social bots**

Les social bots constituent souvent un botnet intelligent qui n'a pas besoin d'être contrôlé par un canal de commande et de contrôle. Il est caractérisé par une autonomie dans son fonctionnement, la faculté de décider à la volée en fonction du contexte, de la mission ou des cibles. Les social bots collectent une grande quantité de données. Après analyse de ces données, ils ont la possibilité de déployer de

2. Une suite de données donnant des motifs dans une séquence.
3. Comme une mémoire cache des réseaux de neurones stockant les informations de l'état précédent

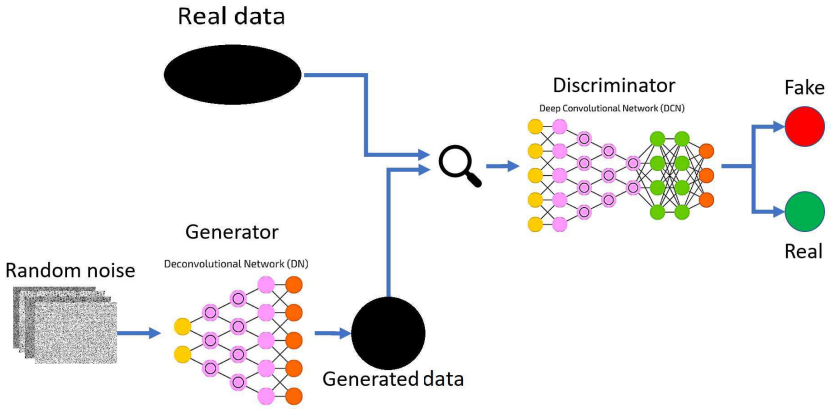


Figure. 1.3 : Modèle du réseau GAN

puissantes campagnes d'influences sur les réseaux sociaux.

Exemple : Social Network Automated Phishing with Reconnaissance (SNAP-R)
Le SNAP-R est un exemple de social bot qui génère des spearphishing de bout en bout de manière automatisée. Ce social bot est caractérisé par l'automatisation du processus de spearphishing pour créer des attaques de spearphishing ciblées à grande échelle avec un taux de réussite élevée. Le SNAP-R classifie en premier une liste de comptes cibles en des groupes dans la phase découverte à l'aide de l'algorithme k-means. Cette classification est souvent basée sur des critères comme les profils publics et le niveau de sociabilité. Dans la phase suivante, celle du spearphishing automatisé, le SNAP-R utilise le NLP pour identifier les centres d'intérêt des cibles. Il utilise ensuite les algorithmes de LSTM et de Markov pour générer du contenu en rapport avec l'historique des cibles. Ce bot est utilisé dans des réseaux sociaux comme twitter avec un grand taux de réussite. La meilleure manière de se défendre contre ce bot est d'utiliser un filtre antispam associé à une sensibilisation massive des utilisateurs.

1.4.3 Processus d'intrusion

Une infiltration est une attaque APT et est utilisée pour collecter des informations sur la cible. C'est une attaque difficile à détecter parce que l'attaquant essaie de garder un comportement normal le plus longtemps possible pour échapper aux mécanismes de surveillance de la sécurité [19]. Une intrusion est souvent effectuée en quatre grandes phases : la reconnaissance, la découverte du système, l'exploitation de vulnérabilités et enfin l'effacement de traces. Le MITRE décrit des intrusions beaucoup plus sophistiquées dans [20] avec plus de détails et plus de phases dans le processus.

1.4.3.1 Reconnaissance

Selon THE-MITREE [20], la reconnaissance est un ensemble de techniques utilisées par un cybercriminel pour collecter activement ou passivement les informations sur la cible. Les informations collectées peuvent servir ultérieurement pour des attaques beaucoup plus sophistiquées. Le cybercriminel peut utiliser des techniques comme le scan de vulnérabilités, le scan de ports, les écoutes clandestines passives, la collecte d'information sur le réseau victime, etc.

1.4.3.2 Découverte du système

Dans cette phase, les intrus utilisent un ensemble de techniques pour observer l'environnement du système dans le but de savoir quoi faire, comment faire et par où commencer. Ils peuvent effectuer notamment des observations sur les comptes, sur les autorisations, sur la structure des fichiers et des dossiers, le sniffing (écoutes clandestines dans le réseau), etc.

1.4.3.3 Exploitation de vulnérabilités

Cette phase est exécutée à partir des résultats de la phase précédente. Dans cette étape, les attaquants exploitent dans la mesure du possible toutes les vulnérabilités trouvées. Le reste et l'ampleur des impacts dépendent des compétences des attaquants.

1.4.3.4 Effacement de traces

Cette phase est la phase de consolidation de l'intrusion. Elle est matérialisée par la réussite dans l'exfiltration de données, la réussite dans le détournement des échanges d'une communication, dans l'interruption ou la destruction des actifs de la victime.

1.4.4 Menace Persistante Avancée

Dans bon nombre de ces exemples, on constatera que les Threat actors cherchent à mettre en œuvre une menace persistante avancée ou Advanced Persistent Threat (APT). Une APT est une menace développée de manière sophistiquée pour permettre à l'attaquant de s'infiltrer dans un système de manière fictive en utilisant plusieurs techniques avancées (reconnaissance, privilege escalation, FootHold, etc.) comme illustré dans la figure 1.4, pour y rester aussi longtemps que possible, jusqu'à ce qu'elle soit identifiée et supprimée. Le but principal d'une APT est le vol de données. Ce qui est alarmant est le temps qu'il faut à une organisation pour déterminer qu'une APT a réussi à s'infiltrer et vit dans le réseau. Un rapport de FireEye [21] nous a montré en 2018 que la moyenne de

jours pour détecter la présence de l'attaquant est de 71 jours en Amérique, 177 jours en Europe, au Moyen-Orient et en Afrique et 204 jours en Asie-Pacifique.

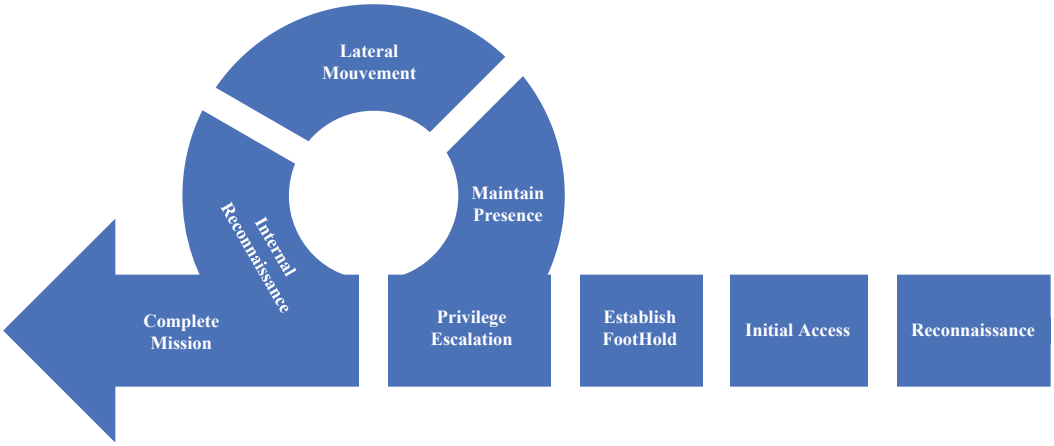


Figure. 1.4 : Menace persistante avancée (APT) par MANDIANT [21]

1.5 Conclusion

Il existe un nombre incalculable d'exemples d'attaques qui se sont produites avec des rapports détaillant ces attaques et leurs impacts. Les attaquants exploitent les vulnérabilités de nos systèmes ou utilisent l'imprudence des utilisateurs par rapport aux logiciels malveillants. Cela montre à quel point nos organisations sont vulnérables. Cette partie illustre clairement les origines des vulnérabilités de notre organisation, mais aussi le grand nombre de possibilités que les attaquants ont à leur disposition pour s'infiltrer dans nos systèmes. Récemment, nous avons noté la naissance des techniques basées sur l'intelligence artificielle utilisées dans les cyberattaques. Ce qui augmente la facilité, la portée ou encore l'impact de ces actions malveillantes. Toute cette malveillance expose continuellement les organisations à de gros risques. Il pourrait être fatal aux organisations, s'ils venaient à échouer sur l'un ou plusieurs objectifs de sécurité. C'est pourquoi tout un ensemble de contre-mesures a été développé pour maintenir le cap sur ces objectifs de sécurité en contrant toute action malveillante ou au pire des cas atténuer leurs effets. Ces contre-mesures sont décidées à l'issue du processus d'analyse de risques. En effet, une analyse permet à nos organisations d'identifier les menaces, d'analyser leurs occurrences et leurs impacts sur les actifs. Une évaluation de risques assigne un niveau ou une valeur aux risques permettant de les classer par ordre d'importance et de les traiter en conséquence. Une fois que les risques sont identifiés et évalués, l'organisation dispose de plusieurs options à travers plusieurs normes (ISO 27002, PCI DSS, NIST, 20 Critical Cybersecurity Controls, Cloud Control's Matrix) pour gérer les risques.

Chapitre 2

Contre-mesures de sécurité et Prévention

2.1 Introduction

Nous avons vu que les cyberattaquants disposent d'une large surface d'attaques. Ils peuvent utiliser les vulnérabilités de notre système ou la négligence des membres de notre organisation. Les attaques peuvent compromettre le fonctionnement ou les données d'un hôte, d'un réseau, de plusieurs réseaux à la fois. Face à ces nombreuses menaces présentes, la plupart de nos organisations mettent en œuvre des politiques de sécurité, déploient des équipements dédiés au sein du réseau ou des applications sur les hôtes. L'ensemble des politiques définies, des mécanismes et les équipements déployés forment ce que l'on appelle les contre-mesures de sécurité. Souvent avant de mettre en place ces contre-mesures, il est d'usage d'effectuer une analyse de risques pour évaluer l'ensemble des menaces présentes, les risques associés et leurs impacts sur notre organisation et décider des contre-mesures adéquates. La plupart de ces contre-mesures sont mises en œuvre par les défenses du périmètre de notre réseau comme les firewalls et les systèmes de détection d'intrusion ou encore par les mécanismes de protection contre les menaces basées sur l'hôte comme les antivirus et les plateformes de sécurité des terminaux.

Les défenses du périmètre de notre réseau nous permettent de le protéger contre tout accès indésirable grâce à des mécanismes et des règles de filtrage. La première défense est le firewall qui agit comme un gardien d'immeuble en contrôlant le flux entrant et sortant de notre système. Le firewall a évolué durant les années en repoussant ses limites, en s'adaptant à l'évolution d'internet et en implémentant davantage de fonctionnalités. Il constitue maintenant un socle où on implémente presque l'ensemble des contre-mesures préventives. Une autre ligne de défense est ce qu'on appelle le système de détection ou de prévention d'intrusion qui est déployé au sein du réseau pour analyser le trafic, détecter

une potentielle intrusion lorsqu'elle se produit et dans certains cas la bloquer. Ce système constitue un socle où on implémente plusieurs contre-mesures comme les audits de sécurité, le suivi des logs, l'analyse des tendances dans les logs et les alarmes, etc. Les mécanismes de protection contre les menaces basées sur l'hôte sont quant à eux utilisés pour renforcer notre sécurité, car en réalité nos équipements terminaux ont toujours été un point d'entrée facile dans notre réseau. Ces mécanismes sécurisent le courriel, la navigation web, l'accès au réseau, les données, etc.

Dans ce chapitre, nous étudions en détail l'ensemble des contre-mesures et les mécanismes de sécurité. La deuxième section est dédiée au processus d'analyse de risques basée sur la norme ISO/IEC 27001. La troisième section est dédiée à l'analyse des contre-mesures disponibles selon les standards internationaux et leur mise en œuvre à l'issue de l'analyse de risques. La quatrième section donne des détails sur les firewalls, les mécanismes de sécurité qu'ils implémentent, et explique leur fonctionnement. La dernière section conclut ce chapitre.

2.2 Analyse et management de risques

Dans cette section, nous discutons de l'importance de l'analyse et du management de risques dans les décisions à prendre concernant la sécurité de notre organisation. Dans le domaine de la sécurité de l'information, le terme vulnérabilité peut aussi être défini comme une absence de contre-mesures ou une insuffisance dans les contre-mesures administrées permettant à une menace de se produire avec une fréquence d'occurrence ou un impact plus important ou encore les deux. Le risque de cette menace est une fonction de la probabilité d'occurrence de celle-ci et son impact sur les actifs de notre organisation. Pour nous protéger de manière efficace et efficiente, nous devons être en mesure d'identifier l'ensemble des menaces qui peuvent compromettre notre sécurité et de mettre en place la défense adéquate contre ces menaces. Des standards comme la série ISO 2700x pour la cybersécurité, la série ETSI EN 303-645 pour les appareils Internet of things a été définie pour aider les organisations dans la sécurisation de leurs ressources.

ISO 27001 est la norme internationale qui définit les spécifications d'un système de management de la sécurité Information Security Management System (ISMS). Un ISMS est l'une des meilleures approches pratiques pour aborder la sécurité de l'information. Il prend en compte les membres, les processus et la technologie de notre organisation. L'évaluation et le management de risques sont au cœur de la norme ISO 27001. Une fois que les risques sont identifiés, nous avons la possibilité de choisir parmi les 35 objectifs de contrôle et les 114 types de contrôle dans 14 catégories proposés par la norme ISO 27002 (voir tableau 2.1).

Tableau 2.1 – Les contrôles proposés par la norme ISO 27002

A.5 Information and security policies	A.8 Asset management	A.12 Operations security	A.15 Supplier relationships
A.6 Organization of information security	A.9 Access Control	A.13 Communications security	A.16 Information security incident management
A.5 Human resources security	A.10 Cryptography	A.14 System Acquisition, development and maintenance	A.17 Information security aspects of business continuity management
A.11 physical and environmental security		A.18 Compliance	

2.2.1 Analyse de risques

Dès lors que de l'argent ou des ressources doivent être dépensés, une analyse de risques doit être effectuée. Une analyse de risque est une technique utilisée pour identifier, évaluer et documenter les facteurs qui peuvent compromettre la réussite dans l'atteinte de nos objectifs de sécurité (Confidentialité, Intégrité et Disponibilité ou CIA). Tout élément matériel ou logiciel comporte des vulnérabilités et donc des menaces associées. On peut modéliser le risque comme étant l'aire d'intersection entre les éléments de ressource, les menaces et les vulnérabilités de ceux-ci. Donc en l'absence d'actif, il n'y a pas de vulnérabilité ni de menace associée. Donc, le recensement des actifs apparait comme une étape importante de l'analyse de risques. Dans l'ensemble des risques, on retrouve les risques physiques, les risques sur les applications, les risques sur le système, les risques sur l'information, les risques sur le réseau, les risques sur le management et les autres risques.

2.2.1.1 Risques physiques

Les risques physiques sont pris en compte par la sécurité de l'information, mais pas par la cybersécurité. Cela concerne les problèmes physiques qui peuvent être liés aux actifs de notre organisation comme :

- Le vol ou la destruction d'équipements ;
- Un problème d'alimentation du système ;
- La panne d'équipements réseau ;
- Les dommages causés par l'homme ou les animaux comme la morsure des rats ou souris sur les câbles ;

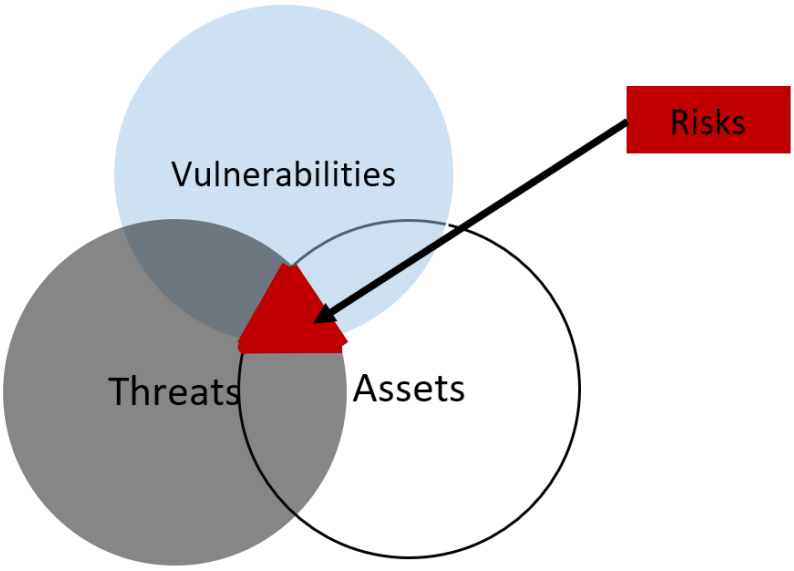


Figure. 2.1 : Les risques en fonction des éléments — vulnérabilités — menaces

- Les radiations électromagnétiques et la vieillesse des équipements, etc.

2.2.1.2 Risques sur l'information

Les résultats de l'exécution d'une menace sur les informations de notre organisation lors d'une transmission ou d'un accès indésirable conduisent généralement à la perte d'accès, à un accès non autorisé, à la modification, à la divulgation de l'information ou encore à la destruction de l'information. Le niveau du risque ici dépend beaucoup de la criticité de l'information.

2.2.1.3 Risques sur le système

La plupart des systèmes d'information sont des systèmes de bases de données. Les mauvaises et les erreurs de configuration de ces systèmes présentent des risques comme les accès indésirables à ces bases de données, les fuites de données, les dénis de services, etc. Ces menaces peuvent avoir un très grand impact sur notre organisation. La sécurité des systèmes englobe la sécurité de la configuration des systèmes de bases de données, la sécurité proprement dite des bases de données par le chiffrement et la sécurité des services du système par l'utilisation de protocoles sécurisés.

2.2.1.4 Risques sur les applications

Les applications présentent de nombreuses vulnérabilités et ces dernières sont très souvent exploitées. Les services réseau (DNS, FTP, Web, Email, etc.) non sécurisés aug-

mentent les risques de pertes de confidentialité et d'intégrité. Le système d'exploitation, les emails et les applications business aussi nécessitent une sécurité adéquate pour minimiser les risques. Les risques sur les applications présentent une probabilité d'occurrence à prendre en compte et leur impact est variable, c'est pourquoi ils sont devenus les principales préoccupations des ingénieurs en sécurité.

2.2.1.5 Risques réseau

Pour l'analyse des risques réseau, nous devons prendre en compte les menaces provenant d'internet, mais aussi celles qui sont internes. Les hackers tentent constamment de s'introduire dans notre réseau pour leur but derrière plusieurs et différentes motivations. Ces accès indésirables peuvent avoir comme impact la perte de données, la divulgation d'informations critiques ainsi que des perturbations dans notre réseau. Cependant, les derniers rapports de cybersécurité soutiennent que les menaces provenant de l'intérieur du réseau sont plus importantes. Les employés amènent maintenant leurs propres au travail ou emportent leur travail à la maison, au restaurant, dans les lieux publics, etc Bring Your Own Device (BYOD). Et ils risquent alors de rapporter des appareils infectés au sein du réseau. La fréquence d'occurrence de ces menaces est très élevée, ce qui augmente le niveau de risque.

2.2.1.6 Risques sur le management

Les principaux risques management sont :

- Une faible sensibilisation à la sécurité parmi les employés ;
- Le vol de données ;
- Des opérations système ne respectant pas les standards comme l'installation de logiciels tiers pouvant comporter des codes malveillants ;
- Des mauvaises règles d'autorisations pouvant ouvrir la porte aux cybercriminels.

L'un des premiers risques d'une organisation du point de vue du management est de ne pas être en conformité avec les politiques de régulations et les lois nationales et internationales concernant la cybersécurité ou la sécurité de l'information de manière générale. Au sein de l'organisation, le faible niveau de conscience sur la sécurité est le principal risque. Moins les membres sont conscients, plus le risque est élevé. La gestion de la sécurité du système est aussi un facteur important de risque. En effet, l'installation tardive de correctifs ou la mauvaise gestion des règles de filtrage sont fréquentes et peuvent aussi ouvrir la porte aux acteurs malveillants. Si l'équipe de management n'est pas suffisamment consciente ou qualifiée, elle n'applique pas des politiques de sécurité strictes, ce qui constitue en effet un risque très important sur la sécurité des informations de notre organisation.

2.2.2 Évaluation et management de risques

Une évaluation de risques nous permet d'identifier les éléments, les vulnérabilités et les menaces dans le but de déterminer le niveau de risques associés et de prendre de bonnes décisions sur les contre-mesures à utiliser. Elle assiste le processus de l'analyse de risques en identifiant les contre-mesures strictement nécessaires à implémenter par notre organisation. L'évaluation de risques a comme avantages [22] :

- L'identification des menaces ;
- La prévention contre les potentielles menaces ;
- La sensibilisation de l'ensemble des membres de l'organisation ;
- La réduction des coûts et dommages en cas de compromissions ;
- La création de politiques et procédures de sécurité efficaces.

Le processus d'évaluation de risques proposé par la norme ISO 27001 se compose de cinq étapes [22] : établir un framework d'évaluation de risques, identifier les risques, analyser les risques, évaluer les risques et sélectionner les contre-mesures parmi les options disponibles.

2.2.2.1 Établissement d'un framework d'évaluation de risques

Cette étape est un sous-ensemble du processus ISMS de l'ISO 27001. Il s'agit d'un ensemble de politiques et de procédures documentées qui définissent une méthodologie formelle d'évaluation des risques. Une méthodologie formelle doit garantir des résultats de l'évaluation des risques cohérents, valides et comparables. Une évaluation des risques doit être cohérente, signifie que deux évaluations des risques effectuées par deux manageurs différents sur le même actif doivent aboutir à la même conclusion sur l'identification des menaces, l'évaluation des vulnérabilités, le niveau de risque appliqué et les contre-mesures sélectionnées par l'organisation. Il convient alors d'une évaluation objective indépendante. Une évaluation des risques doit être valide, signifie que les contre-mesures sélectionnées et les décisions prises ne doivent pas rendre les opérations business essentielles de l'organisation dans leurs cours d'exécution difficiles, voire impossibles. Une évaluation des risques doit être comparable, signifie qu'après avoir effectué une évaluation des risques sur un parc d'ordinateurs portables et sur un parc de serveurs, les contre-mesures probables qui devraient être mises en place sur le serveur devraient être comparables, compte tenu de l'exposition et du risque accrus, à celles mises en place pour le parc d'ordinateurs portables. Les audits par des tiers vérifieront toujours si les évaluations de risques ont été effectuées en respectant ces trois critères.

L'établissement du framework d'évaluation devra commencer par une discussion avec le conseil d'administration de l'organisation sur ce que l'on appelle en anglais risk appetite. Cela mesure ce que l'organisation est prête à perdre par rapport à ce qu'elle est prête

à dépenser pour éviter cette perte. Si ce que l'on peut perdre par exemple en argent, ou en données ou encore en réputation est plus élevé que le prix de la protection alors l'organisation ne va pas hésiter à dépenser pour se protéger. Dans le cas contraire, elle décide souvent de vivre avec ce risque. Il faut aussi choisir la méthode à utiliser pour effectuer l'évaluation. Dans le cas où il faut associer aux risques des chiffres ou des prix pour les quantifier, nous pouvons effectuer une évaluation quantitative de risques. Le calcul (équation 2.1) prend en entrée la probabilité d'occurrence d'un risque basé sur les évidences dans une année (Annualized Rate of Occurrence (ARO)), multiplié aux pertes (d'argent, de temps, etc.) pour une seule occurrence (Single Loss Expectancy). Le résultat de ce calcul donne un facteur de risque qui peut représenter ici les pertes dans une année (Annualized Loss Expectancy (ALE)) [23].

$$\mathcal{ALE} = \mathcal{ARO} * \mathcal{SLE} \tag{2.1}$$

Selon la complexité requise, l'évaluation peut utiliser cette matrice ALE avec plusieurs niveaux de risques et des valeurs numériques (par exemple trois niveaux pour des évaluations moins complexes et cinq niveaux pour des évaluations plus complexes). Cependant, l'impact d'une attaque sur une organisation peut être plus que monétaire (données, accès, blocage, etc.) et dans ces cas-là, il est parfois difficile de donner des valeurs exactes sur les risques. Une évaluation qualitative peut alors nous permettre de voir les points qui ont un niveau de risque élevé dans l'organisation sans avoir de chiffres exacts. Des couleurs où la chaleur de celles-ci révèle la sévérité de l'impact du risque peuvent être associées à la matrice (ou risk heat map, voir tableau 2.2). Le rouge indique une sévérité élevée tandis que l'on a une sévérité moyenne pour le jaune et une sévérité faible pour le vert. Une évaluation qualitative des risques est un moyen très pratique pour une organisation d'identifier rapidement le niveau de risques.

Un autre concept important est d'associer chaque risque identifié à un rôle au sein de l'organisation : c'est ce que l'on appelle risk ownership. Tout risque doit être pris en charge et associé à un responsable, car il est inutile de déterminer qu'il existe un risque alors que personne n'assume la responsabilité de l'application des contre-mesures sélectionnées. Le responsable associé à chaque risque doit être identifié. Ils doivent être documentés dans le registre des risques. Il est souvent recommandé d'associer les risques à des rôles de l'organisation et non à des personnes, car ces personnes peuvent partir. Ceci l'organisation obligerait à réécrire le registre de risques à chaque départ.

2.2.2.2 Identification des risques

Cette étape identifie les risques liés aux personnes, aux processus et à la technologie de l'organisation. Nous devons examiner les habitudes des membres de l'organisation, les

Tableau 2.2 – Risk Matrix/Risk Heat Map [24]

		IMPACT				
		Négligeable 1	Mineur 2	Modéré 3	Majeur 4	Énorme 5
PROBABILITÉ	Certain 5	Modéré 5	Élevé 10	Extrême 15	Extrême 20	Extrême 25
	Probable 4	Modéré 4	Élevé 8	Élevé 12	Extrême 16	Extrême 20
	Possible 3	Faible 3	Modéré 6	Élevé 9	Élevé 12	Extrême 15
	Improbable 2	Faible 2	Modéré 4	Modéré 6	Élevé 8	Élevé 10
	Rare 1	Faible 1	Faible 2	Faible 3	Modéré 4	Modéré 5

faiblesses des processus et les faiblesses de la technologie qui pourraient être exploitées délibérément ou accidentellement par un attaquant. Les évènements naturels doivent aussi être pris en compte (dans le cas de la sécurité de l'information).

Pour commencer, il faut identifier les actifs de l'organisation et donc documenter un registre dédié à ses actifs. Les données et les processus d'affaires font partie des actifs primaires. Il existe aussi un certain nombre d'actifs secondaires susceptibles d'être inclus dans le registre, tels que des matériels, des logiciels, la propriété intellectuelle, etc. Il est nécessaire de déterminer ce qui est essentiel au fonctionnement de l'organisation et de décider des éléments à inclure sur ce registre parmi par exemple les clés USB, les écouteurs, les ordinateurs portables, les sacs pour ordinateur portable, les serveurs, les projecteurs, les écrans, les logiciels, les systèmes d'exploitation, etc. Le registre doit inclure l'ensemble des actifs précieux de l'organisation. L'objectif est de créer un registre des actifs qui identifie et tente d'évaluer les risques sur actifs. Nous devons considérer la valeur de chaque actif en termes de confidentialité, d'intégrité et de disponibilité. Il y a des actifs comme les données circulant au sein de l'organisation, entre sites, sur lesquels la protection de la confidentialité et de l'intégrité sont très importantes. La disponibilité est plus importante sur des actifs comme les serveurs ou encore les ordinateurs avec leurs données et leurs services. La réflexion sur l'état de la triade CIA de chaque actif est une autre étape pour déterminer comment évaluer la valeur de l'actif. Comme avec le concept de risk ownership, chaque actif est géré par un responsable (toujours par rôle) et c'est à ce responsable d'évaluer la valeur de son actif pour l'organisation.

2.2.2.3 Analyse des risques

Nous savons qu'une vulnérabilité est quelque chose de natif à l'actif alors qu'une menace est externe à l'actif. Elle est généralement quelque chose qui peut avoir la possibilité d'exploiter la vulnérabilité. Dans cette troisième étape, les actifs et les risques associés sont déjà identifiés. Nous allons donc essayer de déterminer les décisions les plus pratiques et rentables pour réduire la probabilité d'occurrence de l'exploitation d'une vulnérabilité par une menace. Il est important de déterminer les menaces et vulnérabilités de la manière la plus exacte possible. Par exemple, de nombreuses vulnérabilités logicielles peuvent être déterminées en se référant à des bases de données en ligne où les vulnérabilités seront numérotées et la probabilité peut en être dérivée. Le contrôle approprié à mettre en place peut également être identifié [10]. Des valeurs ou des niveaux d'impact et de probabilité seront attribués à chaque combinaison de vulnérabilité et menace, et cela va être basé sur nos critères prédéterminés. On va donc analyser les risques de toute vulnérabilité menacée sur un actif ou en relation avec celui-ci.

2.2.2.4 Évaluation des risques

À ce stade de notre processus d'évaluation de risques, la démarche à suivre est de comparer les risques identifiés aux seuils de risques acceptables prédéfinis. Cette comparaison a pour but d'identifier les risques les plus urgents et nous aide à définir un ordre de traitement des risques. La comparaison est faite par évaluation quantitative ou qualitative en traçant un graphique d'évaluation de risques (la matrice de risques et le risk heat map, voir tableau 2.2) pour déterminer les risques à traiter en premier. Par exemple, les risques sur le coin supérieur droit sur une matrice de risque doivent toujours être traités en premier. Ainsi, ces graphiques des risques nous permettent de hiérarchiser les risques et de définir l'ordre dans lequel nous les traitons.

2.2.2.5 Sélection de la stratégie de gestion de risques

Maintenant que nous connaissons les risques et l'ordre de priorité, l'étape finale de ce processus est de choisir une stratégie de gestion. L'organisation peut décider :

- D'éviter les risques en mettant en place une politique de prudence et en évitant toute activité à risque ;
- De modifier ou d'atténuer les risques en mettant en place une ou des contre-mesures qui réduisent la probabilité d'occurrence en corrigeant les vulnérabilités par exemple ou l'impact en disposant de plusieurs sauvegardes, ou les deux.
- De transférer les risques à un tiers en souscrivant une assurance risque pour réduire un impact financier

- D'accepter les risques souvent, car le coût des contre-mesures est élevé ou car l'impact est très grand, mais la probabilité est très faible.

2.3 Les contre-mesures de sécurité

Les contre-mesures de sécurité sont les décisions prises et les actions entreprises à l'issue de l'analyse de risques dans le but d'éviter, de transférer, d'accepter les risques identifiés ou encore de les modifier à un niveau acceptable. Les contre-mesures peuvent aussi être séparées en quatre types :

- Les contre-mesures **préventives** qui bloquent les attaques ou fixent les vulnérabilités, ou encore qui tentent d'empêcher l'occurrence de tout incident. Ces contre-mesures se matérialisent souvent par le contrôle d'accès ou encore le chiffrement en mettant en place une sécurité renforcée, par déploiement d'un firewall, des protections, des agents de sécurité, etc. ;
- Les contre-mesures **détectives** en temps réel ou en offline qui tentent de détecter les incidents après qu'ils se soient produits mais n'empêchent pas les incidents de se produire. Celles-ci identifient et enregistrent les tentatives d'intrusion et analysent les tendances dans notre système. Les équipements comme les systèmes de détection d'intrusion lancent des alarmes ou des alertes en cas de détection ;
- Les contre-mesures **correctives** qui tentent d'atténuer l'impact des attaques, de réduire leur sévérité et qui incluent la capacité de se remettre des effets des attaques subies. Les systèmes de prévention d'intrusion par exemple peuvent bloquer les intrusions tandis que les backups peuvent atténuent les effets d'une infection de ransomware ;
- Les contre-mesures **dissuasives** qui n'empêchent pas directement les attaques mais tentent de les rendre très difficiles et de décourager tout individu causeur d'incidents grâce à une très bonne défense. Les technologies comme l'authentification notamment l'authentification multifacteur ou les socles d'empreintes au niveau des portes sont des exemples de dissuasion ;
- Les contre-mesures de **leurre** qui tentent de rendre des réseaux appelées honeypots plus attractives ;
- Les contre-mesures **compensatrices** qui n'empêchent pas les incidents de sécurité. Elles sont utilisées lorsque certains des contre-mesures primaires déjà définies ne sont plus applicables.

Ce qui est intéressant, c'est qu'une technologie pourrait s'adapter à plusieurs de ces types de contre-mesures. Ces contre-mesures de sécurité peuvent être aussi catégorisées par la manière de l'implémentation qui peut être technique, administrative ou encore physique/opérationnelle.

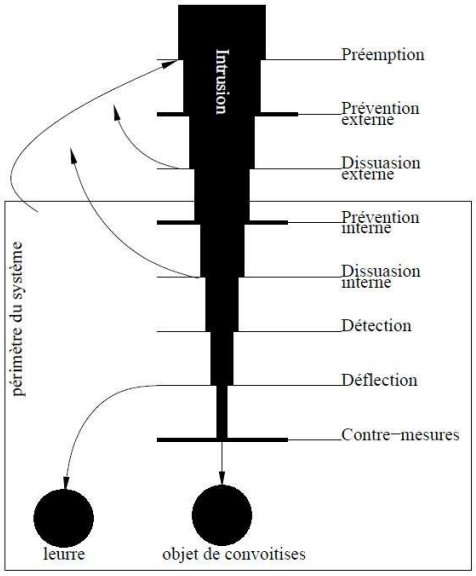


Figure. 2.2 : Les contre-mesures de sécurité [25]

2.3.1 Les contre-mesures administratives

2.3.1.1 L'évaluation de risques, de vulnérabilités et tests de pénétration

Après déploiement, des politiques doivent être définies dans l'optique d'effectuer des évaluations de risques, de vulnérabilités et des tests de pénétration périodiques. Par exemple, les normes comme la famille ISO 2700x fournissent de bonnes pratiques et un calendrier à suivre pour effectuer ces évaluations périodiques. Les tests de pénétration sont très pratiques, car ils permettent de se mettre dans la peau d'un cybercriminel dans ses tentatives d'infiltration.

2.3.1.2 La Formation et la sensibilisation

Une organisation a beau mettre en œuvre des contre-mesures techniques ou peut déployer en son sein des équipements de sécurité. Si ses membres ne sont pas conscients des enjeux de la sécurité, des risques et de leurs impacts, les attaquants parviendront certainement à s'y infiltrer. En effet, la plupart d'entre eux trouvent plus facile d'exploiter la vulnérabilité humaine avec du social engineering que de débaucher beaucoup de temps et autant de ressources pour contourner le firewall. En tenant compte de l'évolution rapide de la technologie, des séminaires de formation périodiques peuvent être décidés ou encore la mise à disposition au sein de l'organisation de manuels ayant le contenu de notre chapitre 1.

Il existe d'autres contre-mesures administratives comme le management des configurations et des changements ou modifications, la planification des situations d'urgence.

2.3.2 Les contre-mesures techniques

Les contre-mesures techniques sont basées sur la technologie. Elles sont souvent illustrées par la présence de deux équipements implémentant ces contre-mesures dans notre système. En effet, les équipements comme les firewalls de nouvelle génération et les Intrusion Prevention System ou Système de Prévention d'Intrusion (IPS)) peuvent implémenter plusieurs technologies comme le chiffrement, le contrôle d'accès, le filtrage web, le filtrage de contenu, la sécurité des emails, etc. L'étude des contre-mesures techniques préventives et détectives est concentrée sur les firewalls et les IDS/IPS.

2.4 Les firewalls de nouvelle génération ou Next Generation Firewall

Les informations et les services disponibles sur internet sont maintenant indispensables à nos organisations et l'accès à internet nous confère plusieurs avantages. Cependant, il permet aussi au monde extérieur d'accéder à notre réseau et d'interagir avec lui. Cette interaction peut être à l'origine d'accès indésirables et d'autres menaces. En effet, les cyberattaquants essayent constamment de pénétrer dans notre réseau. Il y a aussi les logiciels malveillants utilisés pour exploiter les vulnérabilités de notre système. Pour protéger nos systèmes contre ces accès indésirables, le modèle de sécurité d'accès au réseau a été défini. Dans ce modèle, un équipement appelé Firewall est inséré entre notre réseau et internet. Le but de ce déploiement est d'obtenir un nœud dans le réseau où la sécurité et l'audit peuvent être imposés. Tout trafic entrant ou sortant doit passer par le firewall et ceci est réalisé en bloquant physiquement tout accès à notre réseau excepté via ce firewall (figure 2.3). Comme il doit être défini par des politiques de sécurité, seul le trafic autorisé doit traverser le firewall qui doit être quant à lui immuniser contre les attaques. Comme résultat, notre réseau peut alors être protégé contre les attaques provenant d'internet.

De nos jours, il est plus courant de voir notre réseau séparé en différentes zones (internet, intranet, DMZ, local) comme illustré sur la figure 2.4. Cela est rendu facile par les firewalls de nouvelles générations qui incluent la configuration de zones avec leur priorité. La présence dans notre intranet de serveurs accessibles depuis internet présentent de grands risques. La séparation de notre réseau et la migration de ces serveurs vers la zone Demilitarized Zone (DMZ) réduisent ces risques. Avec cette configuration, les internautes n'accéderont pas directement à notre réseau. Par défaut, les firewalls effectuent un contrôle de service en bloquant ou en autorisant des ports. Avec leur évolution, plusieurs contre-mesures techniques sont implémentées sur le firewall et ils sont maintenant capables de contrôler :

- Les utilisateurs et les services qu'ils sont autorisés à utiliser à l'aide de mécanismes

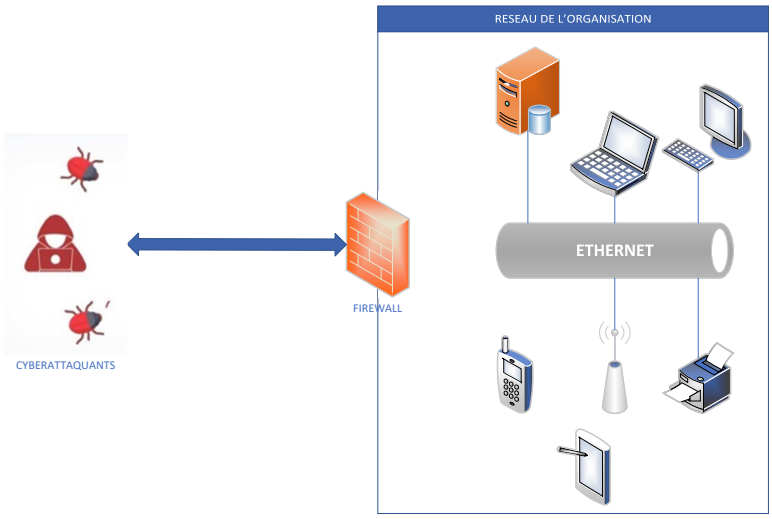


Figure. 2.3 : Le modèle de sécurité d'accès au réseau

d'authentification ;

- La direction dans laquelle une requête pour un service particulier peut être initiée ;
- Le comportement des utilisateurs lors de l'utilisation d'un service.

En plus des fonctionnalités de filtrage de sécurité, le firewall peut inclure la fonctionnalité de surveillance des évènements liés à la sécurité avec une implémentation des audits et des alarmes, une plateforme pratique pour plusieurs fonctions internet non liées à la sécurité telle que la translation d'adresse et la gestion du réseau qui audite ou enregistre l'utilisation d'internet, une plateforme pour le chiffrement de trafic avec l'IPsec VPN entre sites d'une organisation ou encore des fonctionnalités de routage avec l'option d'authentification, le filtrage de contenu [23].

2.4.1 Le chiffrement

Le chiffrement est utilisé dans les échanges pour éviter que toute entité autre que le destinataire ne comprenne le message envoyé. Le but est de rendre le message intelligible que pour les entités ayant en possession la clé utilisée dans le processus de chiffrement. Un algorithme de chiffrement est un ensemble d'opérations mathématiques et logiques utilisées entre le message à envoyer et une clé de chiffrement. Le chiffrement aborde principalement les questions de confidentialité et d'intégrité. Dans le but de permettre une utilisation universelle, les algorithmes de chiffrement sont publiquement connus. Donc la sécurité du chiffrement dépend de la conservation en secret de la clé. Ce qui veut dire que la clé doit être bien conservée et les échanges de clés très sécurisés. Nous avons deux types de chiffrement : le chiffrement symétrique et le chiffrement asymétrique.

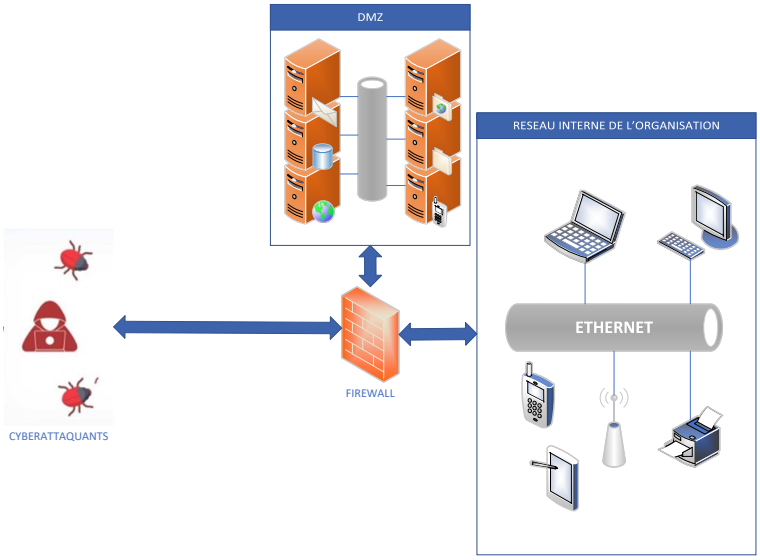


Figure. 2.4 : Les différentes zones d'un réseau

2.4.1.1 Le chiffrement symétrique

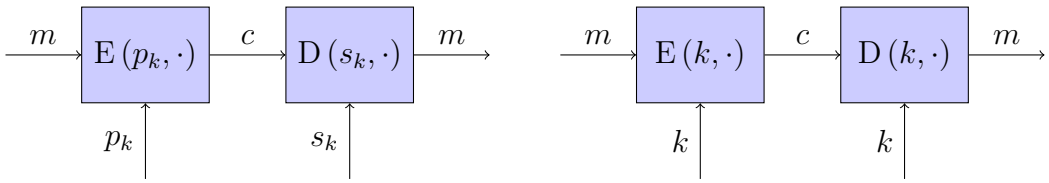
Nous avons deux types d'algorithmes de chiffrement symétrique : le chiffrement par bloc et le chiffrement par flux. Un algorithme de chiffrement symétrique par bloc est constitué d'un ensemble de tours (rounds). Chaque round est à son tour constitué d'opérations de permutation, de substitution, de décalage, de mixage de colonnes. Il y a ajout d'une clé de chiffrement (k) à chaque round. Parmi les algorithmes de chiffrement par bloc, nous avons le Data Encryption Standard (DES) qui fait le chiffrement par bloc de 64 bits en 16 rounds avec une longueur de clé de 56 bits. Nous avons aussi le plus récent Advanced Encryption Standard (AES) qui chiffre par bloc de 128 bits avec un nombre de rounds variable (10, 12 ou 14) selon la longueur de la clé (128 bits, 192 bits ou 256 bits).

Un algorithme de chiffrement par flux est un processus de génération d'une séquence pseudoaléatoire avec la clé de sécurité, suivi d'une opération XOR (octet par octet) entre le message et la séquence générée. L'algorithme Rivest Cipher 4 (RC4) est l'algorithme de chiffrement par flux le plus courant.

2.4.1.2 Le chiffrement asymétrique

C'est un type de chiffrement qui utilise une clé publique (p_k) et une clé privée (s_k). Chaque entité d'une communication possède ces deux clés et partage la clé publique. Lors d'échanges de messages entre entités, l'envoyeur utilise la clé publique du receveur pour chiffrer le message. À la réception, le receveur utilise sa propre clé privée pour déchiffrer le message chiffré. Ceci assure la confidentialité de la communication. En même temps, l'envoyeur peut signer (hachage) le message avec sa clé privé (fingerprint). À la réception,

la clé publique de l'envoyeur est utilisée pour retrouver le message d'origine et le comparer avec le message reçu. C'est un processus de test d'intégrité des échanges. Le chiffrement asymétrique peut aussi être utilisé dans les échanges de clés de session. Étant donné qu'il est plus favorable d'utiliser le chiffrement symétrique pour un grand volume de données, le chiffrement asymétrique est souvent utilisé pour chiffrer la clé symétrique. Pour éviter les attaques par usurpation, on peut utiliser les certificats contenant la clé publique délivrée par l'autorité du certificat (Certificate Authority (CA)). L'algorithme Rivest – Shamir – Adleman (RSA) est le plus utilisé pour le chiffrement asymétrique et l'algorithme Diffie-Helman est le plus utilisé pour chiffrer les clés de session.



(a) Public-Key Architecture **(b)** Symmetric-Key Architecture

Figure. 2.5 : Public-Key vs. Symmetric-Key : Architecture

2.4.2 Le contrôle d'accès - Access Control

Le contrôle d'accès est un ensemble des mécanismes utilisés pour contrôler les appareils, les accès dans le réseau, aux ressources d'une organisation. Le contrôle des appareils permet la détection, l'autorisation et la sécurisation des appareils et médias. Le but est d'authentifier tout utilisateur qui veut accéder au réseau, de déterminer les données qu'il peut accéder et les actions qu'il peut performer. Les éléments dans un processus de contrôle d'accès sont : le demandeur d'accès, le point de contrôle (un serveur) d'accès et un serveur d'authentification (policy server) (voir figure 2.6).

Comme le montre la figure 2.6, le contrôle d'accès peut être effectué par un serveur DHCP, un serveur VLAN ou un point d'accès à travers des méthodes comme le management par DHCP, le contrôle d'accès par VLAN ou encore par la norme IEEE 802.1x au niveau de la couche de liaison de données. La dernière méthode est la solution la plus implémentée. Dans cette dernière solution, le serveur de contrôle d'accès peut intégrer son propre mécanisme d'authentification ou être un client du serveur d'authentification qui va effectuer les authentifications. Le policy server est généralement un serveur Remote Authentication Dial-In User Service (RADIUS). Un serveur RADIUS fournit les fonctionnalités Authentication - Authorization - Accounting (AAA) c.-à-d. d'authentification, d'autorisation et de comptabilité (accounting).

L'authentification est la première phase du processus de contrôle d'accès. Elle permet de vérifier l'identité des demandeurs d'accès. Elle peut être performée par le serveur d'accès

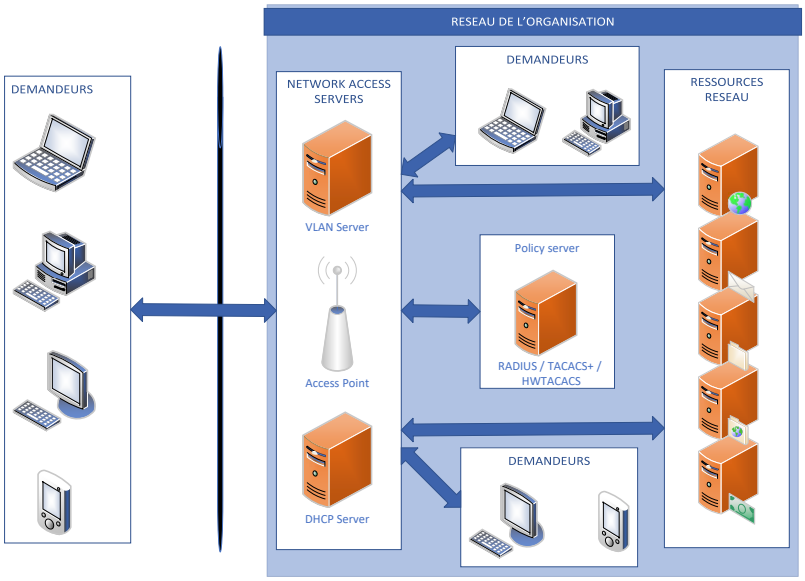


Figure. 2.6 : Le contrôle d'accès au réseau

lui-même ou à travers un serveur RADIUS. Une fois l'utilisateur authentifié, le serveur transmet au serveur d'accès toutes les autorisations comme la qualité de service, les listes de contrôles applicables, les adresses IP allouables, etc. liées à cet utilisateur. Une fois l'utilisateur authentifié et autorisé, le serveur d'accès commence à suivre l'activité de l'utilisateur en notifiant périodiquement le serveur d'authentification.

Avec l'avènement des appareils Internet of Things (IoT) et des BYOD, le contrôle d'accès devient essentiel avec la création de profils pour chaque type d'appareil, segmentant ainsi le réseau. Cela permet de prévenir et de limiter les potentiels dégâts des nouvelles menaces introduites par les IoT et les BYOD.

2.4.3 Le filtrage de contenu

Le filtrage de contenu est une application déployée pour examiner le trafic et déterminer le contenu à bloquer et celui autorisé selon des règles prédéfinies. Il permet de booster la bande passante du réseau d'une organisation en bloquant le contenu vorace en ressources comme les sites de streaming sans rapport avec le travail, et d'augmenter la productivité. En plus du contenu filtré, le filtrage de contenu offre une protection contre les attaques par malwares, les outils d'exploits. On retrouve le filtrage de contenu dans les firewalls, les proxys, les sandbox et les points d'accès WLAN. Il constitue une première ligne de défense contre les attaques web.

Un filtre de contenu peut utiliser une base de données dans laquelle est configurée la liste des sites autorisés ou celle des sites non autorisés. Il peut aussi se baser sur des mots clés ou encore du contenu prédéfini, comme les images avec un caractère particulier. Les

types de filtres de contenu les plus communs sont les filtres web, les filtres internet, les filtres d'engins de recherche, les filtres basés sur DNS, les filtres proxys, les filtres email, les filtres web.

Le filtrage de contenu présente par contre beaucoup de faux positives. En effet, il peut bloquer des sites ou du contenu légitimes et souvent utiles pour le fonctionnement d'une organisation. Des techniques de ML commencent à être intégrées dans les filtres web pour diminuer ces faux positifs. L'engin de recherche de Google intègre déjà des techniques de ML pour améliorer la recherche.

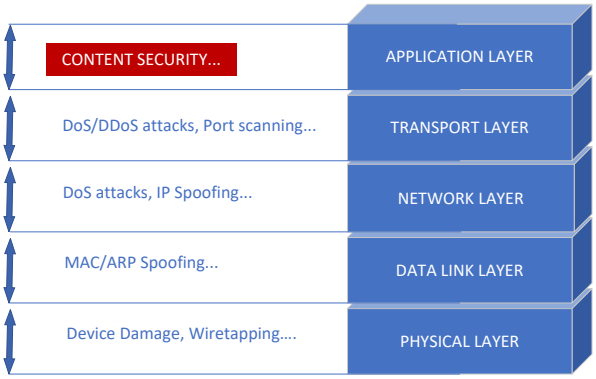
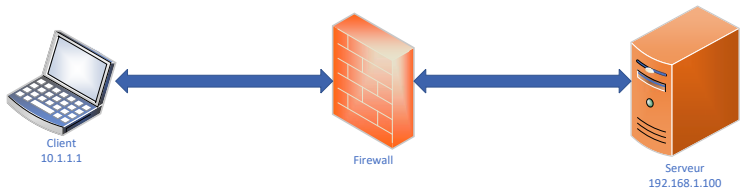


Figure. 2.7 : Le content security dans la pile TCP/IP

2.4.4 Les firewalls sans états et les firewalls avec états

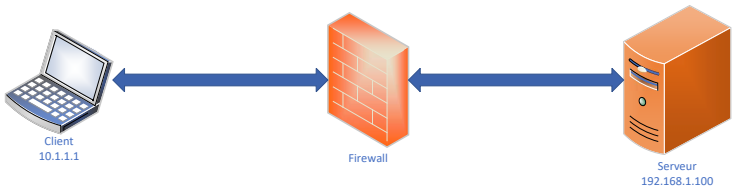
Une communication entre un client et un serveur est souvent basée sur le schéma requête/réponse. Un client web par exemple qui lance une requête s'attend toujours à recevoir une réponse de la part du serveur. Ces échanges sont souvent contrôlés par un firewall. Le premier type de firewall, appelé firewall sans état ou Packet Filter firewall, n'avait aucune idée de ces flux de communication. Ce type de firewall traite chaque paquet individuellement en ne tenant pas compte de l'historique des paquets. Il n'est donc pas en mesure de déterminer le lien entre une requête et une réponse et a besoin d'un ensemble de règles pour chaque direction du trafic pour couvrir une communication dans les deux sens. Dans notre exemple de la Figure 2.8, nous illustrons le comportement du firewall avec sa base des règles qui contient deux règles pour chaque sens. Maintenant, étant donné que le port de la deuxième règle n'est pas fixé et que le firewall n'a aucune idée de l'état de ce flux, un attaquant qui usurpe l'adresse IP du serveur a la possibilité d'envoyer des données au client. Si on suppose que la communication est autorisée (Figure 2.8), les données envoyées par le serveur quel que soit leur caractère (malveillant ou non) vont traverser le firewall.



Rule #	Source IP	Destination IP	Dest Protocol (TCP/UDP)	Dest Port #	Action (Allow/Block)
1	10.1.1.1	192.168.1.100	tcp	80	allow
2	192.168.1.100	10.1.1.1	tcp	any	allow

Figure. 2.8 : Communication à travers un firewall sans état

Avec le même scénario, nous allons maintenant analyser une communication à travers un firewall avec état encore appelé firewall de deuxième génération (figure 2.9). Ce genre de firewall prend en compte les états de communication, c'est pourquoi il n'y a qu'une seule règle par communication dans ses listes de contrôles d'accès.



ACLs

Rule #	Source IP	Destination IP	Dest Protocol (TCP/UDP)	Dest Port #	Action (Allow/Block)
1	10.1.1.1	192.168.1.100	tcp	80	allow

Session Table

Session #	Protocol	Source IP	Source Port	Destination IP	Destination Port
1	tcp	10.1.1.1	15550	192.168.1.100	80

Figure. 2.9 : Communication à travers un firewall avec état

Pour garder la trace de la communication entre le client et le serveur, ce type de firewall prenant en compte le schéma de requête/réponse crée une table de session. Cette table contient des informations sur ce flux particulier comme le numéro de port aléatoire supérieur à 1024 du client qui lance une requête sur le port 80 du serveur. Lorsque le firewall reçoit la réponse du serveur, il vérifie sa liste de contrôle d'accès et constate qu'aucune règle n'autorise ce trafic. Il va alors vérifier l'existence d'une table de session et trouve qu'il y a une session active pour ce trafic. Ce dernier est alors autorisé et transféré vers le client. Et si dans ce même scénario, un cyberattaquant accède au serveur et tente d'envoyer des informations au client sans qu'une session active existe à l'origine, alors le firewall bloque ce flux, car il n'aura pas de correspondance ni dans les bases de règles ni dans la table de session à cause du port TCP de destination différente. C'est ce qui rend le firewall avec état beaucoup plus sécurisé, car il bloque tout paquet n'appartenant à aucune session. En plus, la base des règles est beaucoup moins complexe. C'est le type de firewall le plus couramment utilisé.

2.4.5 Des firewalls de troisième génération et UTM vers les firewalls de nouvelle génération (NGFW)

L'utilisation des protocoles comme HTTP est devenue universelle avec l'explosion du monde web. Ce protocole est utilisé dans diverses applications comme le e-commerce, les applications web, les services cloud, les sites de contenu statique, le streaming, etc. Toutes ces applications utilisent le même port (port 80) et le firewall avec état n'est pas capable de les distinguer. Dans ce cas, les firewalls avec état ne parviennent pas à bloquer les paquets malveillants s'ils utilisent un protocole acceptable. Pour les administrateurs, il est intéressant de faire la distinction entre ces différents contenus pour avoir la possibilité de bloquer ceux qui sont malveillants et de permettre ceux normaux. Dans ce concept de content security, le firewall doit inspecter les paquets jusqu'à la charge utile (data payload) pour comprendre le fonctionnement de certains protocoles comme HTTP.

La 3e génération de firewall inclut cette fonctionnalité de filtrage au niveau applicatif. Ces firewalls comprennent les protocoles au niveau applicatifs comme HTTP, FTP, DNS, etc. et contrôlent les différentes utilisations d'un même protocole. Dans le cas du protocole HTTP, ils peuvent différencier le trafic de navigation dans un site, d'un site web de partage de fichiers, d'un site web d'e-commerce ou encore d'un réseau social. Les menaces apparaissent de plus en plus fréquemment, les attaquants sont de plus en plus nombreux et habiles. Diverses solutions ont été créées pour adresser spécifiquement chaque menace. Les antivirus ont été développés pour contrer les virus, le firewall a été conçu pour bloquer les attaques réseau, les serveurs proxys ont été conçus pour contrôler la navigation web avec des filtres d'URL. Ces solutions sont certes efficaces, mais présentent l'inconvénient d'être complexe d'un point de vue de la maintenance. La gestion de ces solutions séparément est un travail ardu pour les administrateurs. Dans le but de faciliter la tâche à ces administrateurs, l'équipement Unified Threat Management (Unified Threat Management (UTM)) a été conçu. Cet équipement regroupe plusieurs solutions de sécurité tout en simplifiant le management de celles-ci. Un UTM peut être composé d'un firewall, d'un filtrage de contenu, d'un filtrage d'URL, d'un antivirus, d'une protection antispam, d'une protection anti-DDoS.

Pour suivre le paysage des menaces en constante évolution, les Next Generation Firewall (NGFW) sont créés pour gérer la grande surface d'attaque augmentée par l'utilisation des services cloud et des appareils mobiles et IoT. Les cybercriminels utilisent maintenant des méthodes multiples et plus sophistiquées, les attaques peuvent venir de partout. NGFW doit empêcher les cyberattaques tout en assurant la sécurité, la fiabilité et les performances du réseau à tous les bords du réseau. Le NGFW a plusieurs points de contrôle de sécurité. Lorsqu'un paquet arrive, le NGFW procède d'abord à un filtrage de paquets, si le paquet est autorisé à passer, la deuxième étape est une inspection ap-

profondie des paquets si nécessaire. L'inspection approfondie des paquets (deep packet inspection en anglais) est une inspection de la charge utile des données des paquets. Il peut être requis pour un type spécifique de paquets, un type spécifique d'adresses IP par exemple. Si l'inspection approfondie des paquets détecte quelque chose de suspect, elle l'envoie au sandbox pour un examen plus approfondi. NGFW implémente la sécurité au niveau de l'application (également appelée content security) par utilisateur, par comportement. Le NGFW a battu l'UTM par ses capacités d'inspection haute performance et une plus grande visibilité du réseau tout en incluant les mêmes solutions de sécurité. L'inspection de haute performance inclut les applications, les ressources du système, les données cryptées qui traversent le réseau et le stockage de données, etc.

2.4.6 Les Web Application Firewall

Un Web Application Firewall (WAF) est le firewall utilisé dans les applications web. Le WAF fournit une autre couche de protection pour l'application en plus du firewall réseau déjà déployé. Il est spécialement conçu pour protéger l'application web hébergée dans un serveur. Il suit et applique des règles de filtrage aux échanges HTTP/HTTPS. Il cible le contenu d'une application web spécifique, au niveau application OSI. Il peut aussi permettre ou bloquer un trafic en se basant sur les entrées de requêtes attendues. La plupart des exploits utilisent des méthodes avec des entrées de requêtes inattendues par le serveur. Le WAF est très efficace contre les attaques de type injection de SQL, de type Cross-Site Scripting, de type inclusion de fichiers et les erreurs de configuration de la sécurité. Les premières générations de WAF utilisent des listes de blocage et des attributs HTTP basés sur la signature pour alerter le firewall réseau d'une attaque. Avec le temps, la technique basée sur les signatures est devenue moins efficace et génère beaucoup de fausses alarmes, augmentant ainsi la charge de travail de l'équipe chargée de la sécurité. Les dernières générations de WAF sont devenues plus intelligentes avec l'inclusion de la capacité d'apprentissage de comportement. Le WAF peut créer un profil d'utilisation et de comportement d'une application pour ensuite évaluer les comportements futurs puis les classer comme normales ou anormales donc suspects. Les WAF incluent maintenant le monitoring de session et l'analyse heuristique permettant au firewall de détecter les malwares polymorphes. Le WAF peut être un équipement autonome ou un logiciel placé entre le client et le serveur. Malgré toutes ces fonctionnalités, et avec l'explosion du nombre de techniques et de mutations de méthodes existantes ou de nouveaux exploits, les WAF ont montré leurs limites et sont de plus en plus incapables de détecter les attaques zero-day. C'est pourquoi la technologie s'est tournée vers l'intelligence artificielle supervisée par l'homme pour analyser les comportements des usagers à la vitesse des machines tout en s'adaptant au changement des attributs des menaces.

2.4.7 La sécurité des terminaux – Endpoint Security

Le terminal est l'appareil utilisé par l'utilisateur final. Il peut être un ordinateur, un laptop, un smartphone et plus récemment un appareil IoT. Les terminaux ont toujours été un point d'entrée facile vers le réseau. «Pourquoi essayer de contourner un pare-feu si je peux exploiter un utilisateur négligent et crédule grâce au social engineering ?» : se demanderait un attaquant. En plus de cela, le nombre d'appareils et de connexions en ligne augmentent fortement. Ce qui augmente et diversifie les vecteurs d'attaques. Dans le cas d'un réseau non connecté à internet, les attaquants, les insiders par exemple, utilisent souvent des supports de stockage amovibles comme les clés USB. Dans cet exemple, la défense appropriée peut être l'utilisation d'un produit de sécurité pour les terminaux.

2.4.7.1 Les antivirus – les anti-malwares

L'antivirus est un logiciel qui scanne les appareils et les disques durs dans le but de trouver la présence potentielle d'un malware. Les antivirus se basent sur la signature et recherchent des caractéristiques spécifiques c.-à-d. la signature d'un malware. Les actions de l'antivirus à la détection d'un malware sont soit la mise en quarantaine, soit la suppression. À cause d'internet, les vecteurs d'attaques se sont multipliés (email phishing, sites web infectés, les réseaux sociaux, le «BYOD» qui permet aux employés d'utiliser leurs propres appareils au travail, etc.). Récemment, les malwares ont commencé à acquérir le caractère de mutant des virus biologiques. Ces malwares appelés malwares polymorphes sont conçus pour muter tous seuls. Ce qui donne naissance à des familles de malwares polymorphes et rend la détection par signature inefficace.

2.4.7.2 Les plateformes de protections des terminaux – EPP

Ce deuxième produit pour la sécurité des terminaux a été conçu pour contrer les attaques basées sur les fichiers infectés, les malwares polymorphes, etc. Il arrive de retrouver du code malveillant dans un fichier qui s'exécute une fois le fichier ouvert. Cette plateforme implémente d'autres contrôles préventifs dans le but de stopper le malware avant que le terminal ne soit infecté. Les EPP fournissent plusieurs services qui sont focalisés sur la prévention comme les antivirus, les firewalls hôtes, le filtrage web, le chiffrement des données, et le contrôle des appareils. Cependant, aucun de ces mécanismes n'est infaillible. Ils ne peuvent pas assurer la prévention à 100 % à cause des méthodes d'attaques de plus en plus sophistiquées et une surface de plus en plus large.

2.4.7.3 Les outils de détection et de réponse pour les terminaux – EDR

Il est quasi impossible d'assurer une prévention totale, c'est pourquoi le troisième produit de sécurité, c.-à-d. l'outil de détection et de réponse pour les terminaux (EndPoint Detection Response (EDR)), a été développé pour détecter, enquêter et bloquer les activités suspicieuses sur les terminaux. Il fournit aux analystes de sécurité (SOC) des informations Threat intelligence et un ensemble d'outils, dont un outil de recherche de forensique, pour analyser les attaques et identifier les Indicators of Compromission (IoC). Grâce aux EDR, les analystes sont capables de détecter certains malwares qui peuvent rester indétectables durant des mois voire des années à l'intérieur d'un réseau. Avec la capacité d'enquêter sur une attaque et de connaitre son anatomie, l'outil peut aussi détecter une attaque en temps réel. Des modules de remédiation ont été ajoutés pour permettre aux analystes de recueillir plus d'informations sur les terminaux, de bloquer des adresses IP spécifiques, de bannir certains processus ou encore d'isoler certains terminaux.

Les premières générations d'EDR s'utilisaient de manière manuelle, ce qui prenait beaucoup de temps aux administrateurs. En plus, ces EDR sont inefficaces contre les attaques rapides comme les ransomwares. Ces outils présentaient beaucoup de limites telles que :

- Une incompatibilité avec les autres applications de sécurité qui entravaient son habilité à répondre efficacement et au moment opportun ;
- Une très grande expertise requise pour leur configuration et leur utilisation ;
- Un très grand nombre de fausses alarmes.

Les deuxièmes générations d'EDR corrigent ces problèmes. Leur conception est basée sur des politiques de sécurité. Ces EDR sont automatisées et donnent aux analystes la possibilité de leur apprendre à répondre de manière spécifique et proactive. Ils peuvent détecter un programme ou un script qui présente un comportement anormal. Les activités malicieuses déclenchent des blocages automatiques pour éviter l'exfiltration de données ou encore les intrusions. Ils peuvent aussi stopper les attaques ransomwares en temps réel sans nécessairement retirer le périphérique ou perturber le fonctionnement des opérations de notre organisation. En plus, ces outils peuvent ajouter des patchs virtuels et des politiques qui appliquent des restrictions jusqu'à la correction d'une vulnérabilité découverte.

Nous avons enfin les XDR qui sont des extensions des EDR+EPP (voir figure 2.10). Ces derniers sont encore plus performants, et commencent à inclure des techniques de Machine Learning. Ils seront à l'avenir capables de prévenir, de détecter les malwares et d'y remédier à la vitesse de la machine. Les XDR gèrent aussi les couches Access et network du cloud.

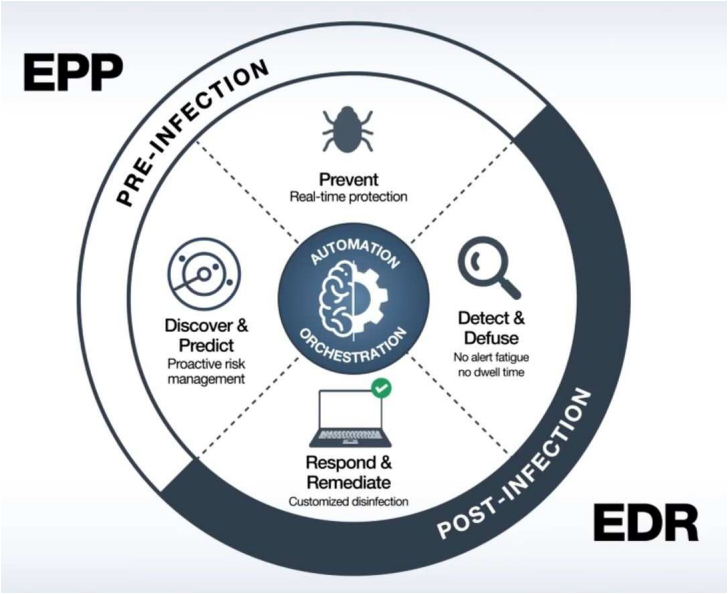


Figure. 2.10 : Les produits de sécurité pour terminaux [16]

2.4.8 La sécurité du cloud

Dans le temps, la plupart des organisations hébergeaient leurs applications au sein de leurs locaux, dans leurs propres datacenters. Cela leur permettait d'appliquer les mécanismes de sécurité appropriés pour gérer des problèmes de sécurité spécifiques. Cependant, cette pratique est coûteuse et présente beaucoup d'inconvénients surtout pour les startups. De nos jours, les organisations utilisent de plus en plus les services cloud comme Amazon Web Services (AWS), Alibaba Cloud, Google Cloud, Microsoft Azure, IBM cloud, etc. En utilisant le Cloud, les organisations assurent une plus grande disponibilité pour leurs ressources et leurs applications. Par contre, même si la plupart des fournisseurs cloud proposent quelques outils de sécurité qui aident à sécuriser les applications et les données des clients, ces outils assurent uniquement des fonctions basiques de sécurité. Les fournisseurs cloud et les clients se doivent alors de partager les responsabilités. Les fournisseurs assurent la disponibilité pendant que les clients assurent la confidentialité, l'intégrité, l'authentification. Il y a des organisations qui fonctionnent en hybride où certaines de leurs applications sont chez le fournisseur cloud, d'autres hébergées en local.

2.5 Conclusion

L'analyse de risques est une étape importante dans la mise en oeuvre d'une sécurité digne de ce nom au sein d'une organisation. En passant par une identification des menaces, par leurs analyses et leurs évaluations, une organisation conduisant l'analyse a toutes les informations nécessaires propres à sa structure pour décider des contre-mesures adéquates. Grâce à des frameworks comme la norme ISO 2700x, une analyse de risques

permet aussi aux organisations de ne pas dépenser de l'argent inutilement. Après cette première étape, une organisation entre dans la phase de mise en place et de déploiement. Il existe une variété de contre-mesures administratives, techniques, etc. à disposition pour mettre en oeuvre les politiques de sécurité définies. Les politiques préventives sont toujours nécessaires. Des équipements comme les firewalls représentent des socles pour les contre-mesures techniques préventives. En effet, le Firewall implémente la plupart des contre-mesures techniques citées dans le tableau 2.1. Il est toujours nécessaire de former les membres de l'organisation à la sécurité et d'effectuer des évaluations de vulnérabilités et tests de pénétration périodiques pour assurer une sécurité pérenne et adaptative. Malgré la très bonne efficacité de la prévention, il est impossible de l'assurer totalement. c'est pourquoi des techniques comme les EDR ont été développées. Il est alors nécessaire de penser à mettre en œuvre des contre-mesures détectives, correctives qui sont essentielles contre la sophistication des attaques et l'avènement des attaques zero-day.

Chapitre 3

Systèmes de détection/prévention d'intrusion

3.1 Introduction

Face aux nombreuses menaces, nous disposons d'une panoplie de contre-mesures techniques pour prévenir toute attaque ou intrusion dans nos organisations. De nos jours, les organisations définissent des politiques de sécurité, entraînent leurs membres, déploient des équipements et technologies pour contrer les attaquants. Et pourtant malgré toutes ces contre-mesures préventives, nos systèmes reste toujours vulnérables et sont constamment sous les menaces d'attaques notamment les attaques zero-day. En effet, il existe des failles de sécurité encore inconnues au sein de nos équipements, de nos applications, de nos configurations que les cybercriminels peuvent découvrir et exploiter avant leur correction. Pour contrer les attaques, des équipements comme les systèmes de détection d'intrusion (IDS) et les Endpoint Detection and Response (EDR/XDR) sont déployées. Ces technologies offrent une deuxième ligne de défense à nos organisations.

Le challenge actuel des fournisseurs et des chercheurs est de découvrir et de corriger ces vulnérabilités avant qu'elles ne soient exploitées par les attaquants. Des programmes comme les Bug Bounty où des chercheurs soumettent des failles de sécurité découvertes pour une somme d'argent sont mis en place. Il y a aussi des frameworks proposés pour détecter l'exploitation d'une vulnérabilité 0-day dans la littérature. Les fournisseurs d'équipements quant à eux utilisent ce que l'on appelle la Threat Intelligence pour partager les renseignements entre eux et pour mieux se défendre contre ce genre d'attaques aux impacts encore inconnus. En même temps, les chercheurs en cybersécurité se sont orientés vers l'IA pour trouver le moyen de détecter les attaques inconnues mais aussi d'effectuer les traitements à la vitesse de la machine pour répondre aux attaques rapides.

Dans ce chapitre, nous étudions les systèmes IDS/IPS avec toutes leurs fonction-

nalités, leurs limites et les tendances de la cybersécurité. Dans la deuxième section, les systèmes de détection d'intrusion seront abordés. Une analyse de la situation actuelle sur la course à la découverte de failles entre la communauté de cybersécurité (chercheurs et fournisseurs) et les cyberattaquants, est faite dans la troisième section. Dans la quatrième section, nous étudions les stratégies de détection basées sur l'IA avant de conclure ce chapitre.

3.2 Système de détection/prévention d'intrusion

Dans le chapitre précédent, nous avons étudié un ensemble de contre-mesures préventives symbolisées par la définition de politiques de sécurité et le déploiement de firewalls et de produits de sécurité pour terminaux. Ce qui remplit efficacement le principe de Hardening qui signifie rendre difficile l'intrusion. Cependant, il est impossible de prévenir toutes les intrusions. C'est pourquoi une organisation se doit de définir en parallèle des contre-mesures détectives et correctives. Un équipement qui symbolise ce type de politiques est le système de détection d'intrusion (IDS) accompagné d'une gestion des informations et des évènements de sécurité (SIEM). La détection d'intrusion est un ensemble de mécanismes et de pratiques utilisés pour détecter toute activité pouvant entraîner une défaillance de sécurité et diagnostiquer les intrusions et les attaques. Un IDS est alors une implémentation des mécanismes et pratiques de la détection d'intrusion. En effet, un IDS, à l'aide de ses sondes, analyse l'ensemble du réseau d'une organisation dans le but de détecter les intrusions. Lorsqu'un évènement survient, l'IDS alerte ou alarme l'administrateur ou l'équipe NOC/SOC [1] qui vont déployer des mécanismes de réponse à cet évènement.

3.2.1 Taxonomie des IDS/IPS

Dans cette section, nous allons définir les IDS/IPS et leurs fonctionnalités et décrire leur fonctionnement. Comme ils sont définis, dans la section qui précède, l'IDS est l'implémentation d'un ensemble de mécanismes, d'opérations, d'actions avant, pendant ou après la détection d'une potentielle intrusion. Les données analysées sont issues du trafic réseau, des logs du système, des logs du firewall et des logs d'applications. Le type de données analysées et le type d'IDS dépendent en particulier de la plateforme de déploiement. Il y a principalement deux catégories d'IDS : l'IDS réseau (NIDS) et l'IDS hôte (HIDS) [1]. La figure 3.1 donne une taxonomie complète des IDS.

1. Network Operation Center/Security Operation Center

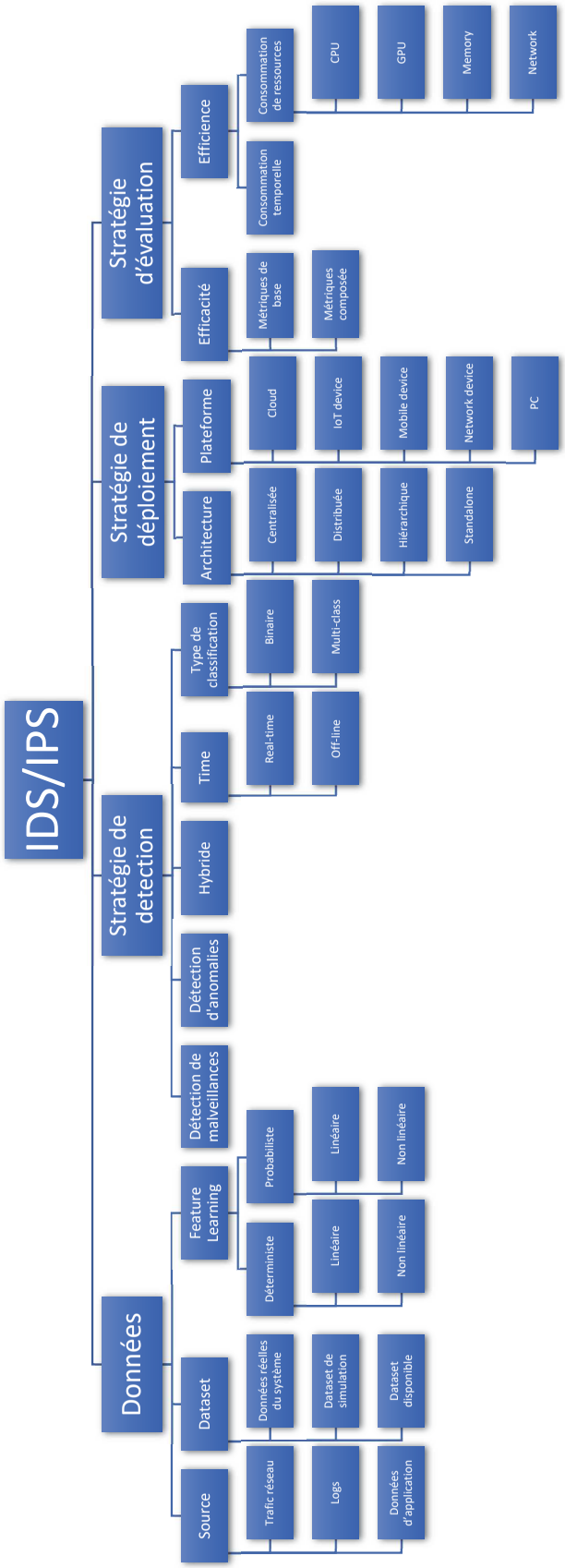


Figure. 3.1 : La taxonomie des IDS

3.2.1.1 Fonctionnement des NIDS/NIPS

Un NIDS est un équipement dédié qui analyse constamment le trafic du réseau de la couche 1 à la couche 7 du modèle OSI à l'aide de sondes pour déterminer l'existence potentielle d'intrusion. Son architecture illustré dans la figure 3.2 comprend quatre modules [26] :

- Le module de collecte qui est composé de collecteurs. Ils capturent les paquets et envoient des copies aux autres modules pour le besoin d'analyses et de détection ;
- Le module d'extraction de données qui utilisent des outils et des techniques d'analyse de données pour extraire les caractéristiques importantes pour la détection d'intrusion ;
- Le module d'analyse et de détection d'intrusions qui compare les données reçues aux signatures aux profils ou autres stockés dans une base de connaissances. Le résultat de la comparaison détermine s'il y a intrusion ;
- Le module de réponse qui reçoit le résultat de la comparaison et déclenche une réponse si un évènement a lieu. Les différents types de réponses seront discutés dans la section des actions des NIDS/NIPS.

Dans le module d'analyse et de détection, il y a souvent un moteur de reconstruction de fragments où les fragments détectés seront envoyés pour être analyser après reconstruction.

Cet équipement a deux modes de fonctionnements. Il y a le mode promiscuité où le NIDS possède un ou plusieurs interfaces réseau pour collecter des copies du trafic sur

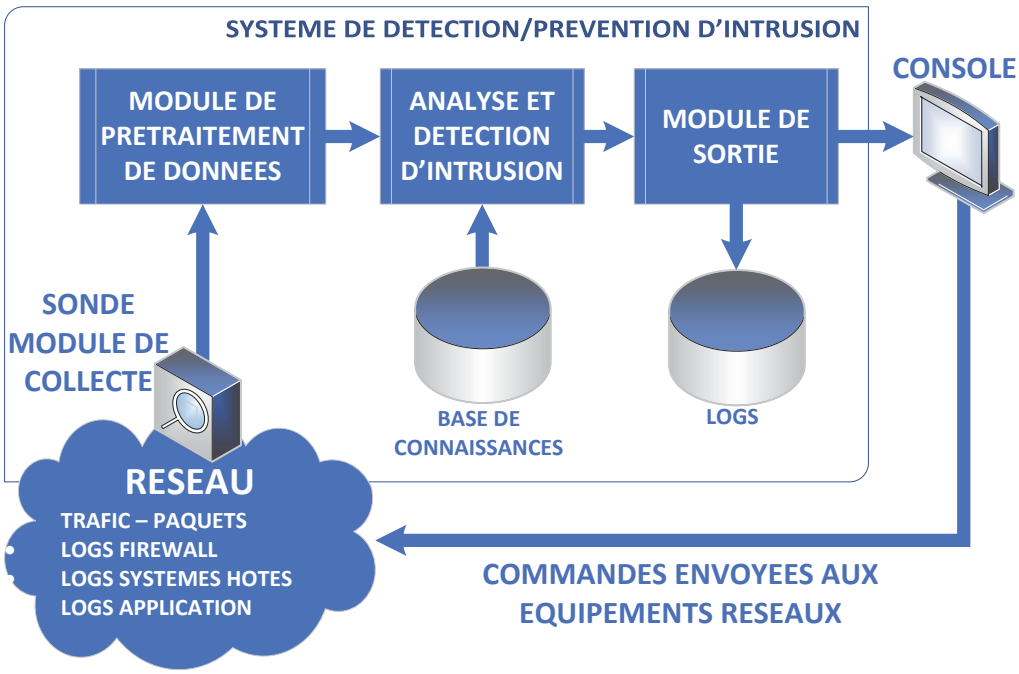


Figure. 3.2 : L'architecture des NIDS

lesquelles des traitements et analyses seront effectués dans le but de détecter les activités malicieuses et de répondre en envoyant des commandes de blocage aux switches, aux routeurs ou le plus souvent aux firewalls. Il y a aussi le mode furtif où le NIDS est présent dans le réseau sans avoir d'adresse IP et sans être attaché à aucune pile de protocoles. Le NIPS, quant à lui qui est en mode inline, analyse le trafic en temps réel et agit directement sur celui-ci en cas d'évènement afin d'empêcher l'intrusion d'aller plus loin. Les paquets ne traversent le NIPS que s'ils sont clean après une analyse approfondie. Il faut noter que les IDS et les IPS ne sont différents que par la position de la sonde et la nature de leur réponse à une intrusion, mais fonctionnent de la même manière.

3.2.1.2 Emplacement des sondes

La mise en œuvre de la sécurité dans une organisation dépend des politiques de sécurité définies. Dans la plupart des cas, l'organisation met en oeuvre des politiques préventives et d'autres politiques, après l'occurrence d'un évènement, comme la détection et la correction. Dans notre contexte, l'emplacement des collecteurs d'un IDS déployé dépend des politiques et des zones critiques du réseau de celle-ci. Néanmoins, il est courant de placer les sondes comme illustré sur la figure 3.4 :

- Dans la zone interne du réseau pour détecter et éventuellement contrer les intrusions depuis et vers l'intérieur ;
- Dans la zone DMZ au cas où les serveurs publics de l'organisation soient attaqués ou contrôlés par un cybercriminel ;

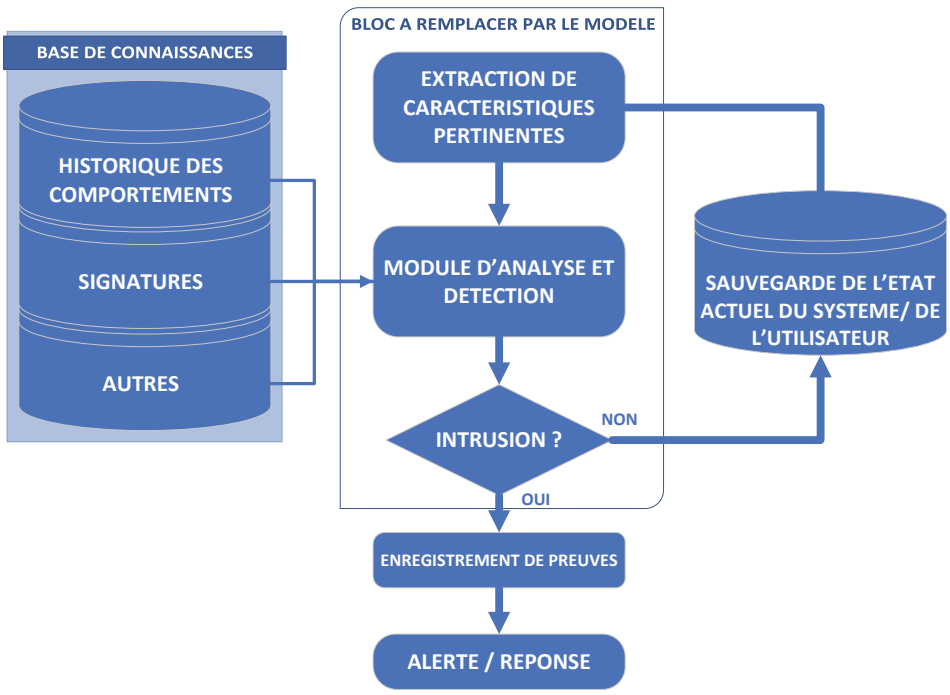


Figure. 3.3 : Principe de détection des NIDS actuels

- Dans la zone externe du réseau ou sur le firewall pour analyser les trafics entrant et sortant. Ce dernier emplacement n'est pas très pratique, car dans celui-ci, l'IDS génère trop d'alarmes donc trop de faux positifs.

Il faut noter qu'il y a deux options de connexions des sondes aux réseaux. La première option est d'utiliser le port mirroring pour que le switch, le routeur ou le firewall envoie une copie de tout trafic le traversant. Cette option de connexion est en phase avec le mode promiscuité. La deuxième option est de brancher directement un Test Access Port (TAP) sur le segment du réseau. La première option a pour avantages d'être simple à configurer et peu coûteuse, mais n'est pas très efficace, car il y a des pertes de paquets et une latence introduite. La deuxième option nécessite un équipement TAP, mais assure une fiabilité dans les mécanismes de copie de flux et n'introduit pas de latence.

3.2.1.3 Actions des NIDS/NIPS

Comme nous l'avons mentionné plus haut, les IDS diffèrent des IPS par leurs actions après détection. Lorsqu'une intrusion est détectée, un IDS peut performer les actions suivantes :

- Mettre en logs les détails de l'intrusion comme les adresses attaquant et cible, le protocole utilisé, la charge utile, le timestamp, etc. ;
- Sauvegarder tous les paquets ou uniquement les paquets ayant déclenché les alarmes ;
- Notifier l'intrusion sur l'écran de la console de gestion ;

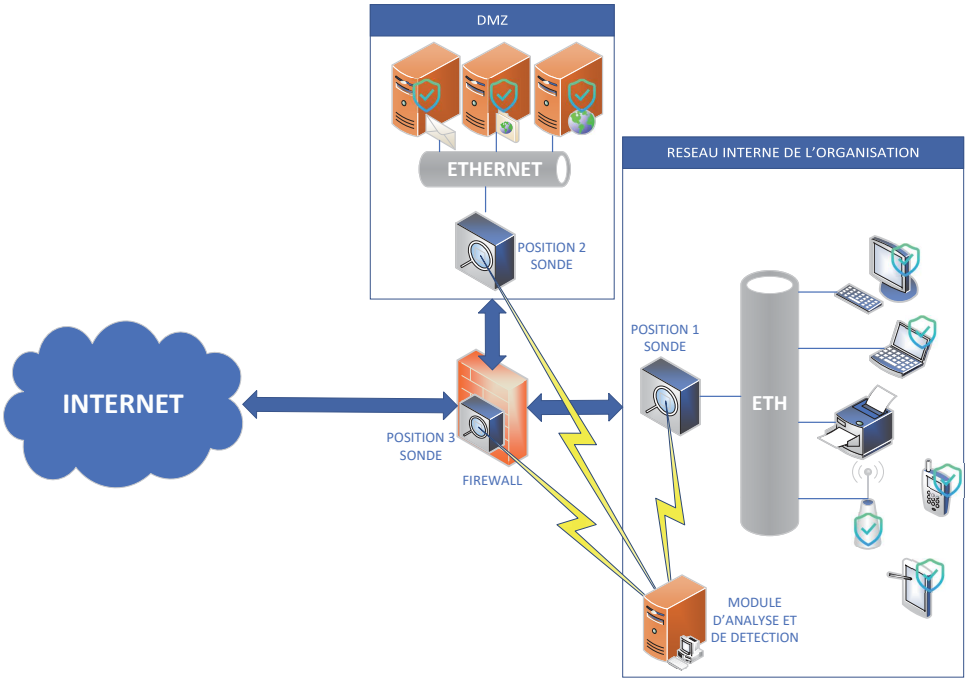


Figure. 3.4 : Emplacement des sondes NIDS

- Notifier l'intrusion à l'administrateur par mail ou par trap SNMP pour permettre à l'administrateur ou à l'équipe NOC/SOC au poste d'être au courant à temps ;
- Démarrer une autre application pour déclencher l'alerte (alertes auditives ou par SMS, etc.)

Lorsqu'une intrusion est détectée en présence d'un IPS, il y a deux possibilités de réponses : une réponse active ou une réponse passive.

- Une réponse passive de l'IPS tente de bloquer l'intrusion tout en le maintenant furtif. Dans ce cas, l'IPS a comme action l'envoi d'ordres aux firewalls, aux routeurs et autres équipements pour bloquer l'intrusion.
- Par contre si un IPS a une réponse active, il peut renvoyer des segments RST ou FIN pour couper rapidement toute connexion TCP suspecte. Ce qui le rend détectable et exposé à une contre-offensive de la part de l'attaquant. Il peut aussi actionner une reconfiguration d'équipements qui peut être immédiate pour bloquer l'intrusion.

3.2.1.4 HIDS/HIPS

Comme pour les contre-mesures préventives avec les produits de sécurité pour terminaux, les IDS aussi ont leur version hôte (HIDS) déployée sur un système hôte spécifique. Contrairement au NIDS, le HIDS surveille uniquement l'hôte dans lequel il est installé. Il fonctionne comme un service standard et utilise les ressources de l'hôte pour son fonctionnement. Il peut effectuer des actions comme :

- L'analyse de logs ;
- Le contrôle d'accès au système ;
- La vérification de l'intégrité des systèmes de fichiers ;
- L'accès en profondeur aux composants système comme les registres ;
- La surveillance de l'OS et des applications en local ;
- La capture de paquets entrant/sortant de l'hôte pour déceler une intrusion.

Il est souvent utilisé en complément aux autres solutions de sécurité sur des hôtes prioritaires qui ont un fonctionnement critique pour l'organisation.

3.2.1.5 Exemple d'IDS : SNORT NIDS

Il existe plusieurs solutions d'IDS dans le marché. Nous avons les IDS/IPS des fournisseurs comme Cisco, Huawei, Juniper, etc. qui sont des équipements dédiés, les NIDS open source comme SNORT, Surricata, les HIDS open source comme Tripwire, OSSEC, AIDE. En guise d'exemple, nous allons étudier SNORT.

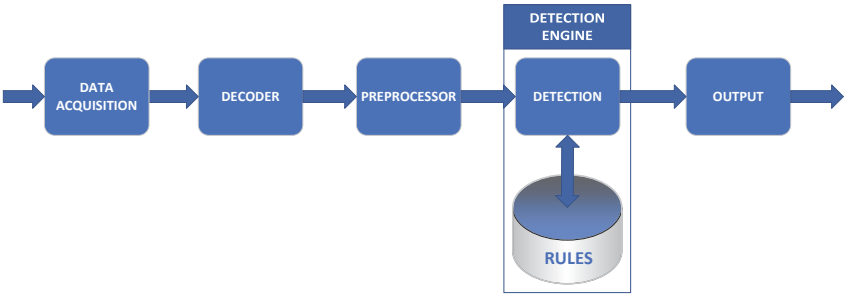


Figure. 3.5 : Schéma de fonctionnement de SNORT

SNORT est une solution logicielle open source (SNORT 2) qui peut fonctionner comme un sniffer, un Packet Logger ou NIDS/NIPS. Il peut fonctionner en mode inline où il a la capacité de bloquer, en mode passive (TAP) et en mode inline-test où il notifie des paquets à bloquer. Le mode NIDS/NIPS nous intéresse particulièrement. Depuis sa version 2.9 [27], les données à analyser sont recueillies grâce aux librairies d'acquisition de données (DAQ). Le DAQ peut être en mode PCAP pour recueillir les paquets capturés par un sniffer comme Wireshark. Il peut aussi être en mode NFQ/IFQ pour les paquets iptables, en mode IPFW pour les systèmes BSD [2]. A l'acquisition des données, SNORT décode les paquets (Ethernet, IP, TCP) en vérifiant la présence d'erreurs et d'anomalies dans les entêtes. Il fait ensuite appel aux plugins pré-processeurs avant d'exécuter le code de l'engin de détection pour enfin fournir un verdict et logger les évènements. Le rôle d'un pré-processeur est de gérer un type de trafic spécifique, de tenter d'éliminer les potentielles tentatives d'évasion d'IDS. Ces plugins gèrent la fragmentation IP et le réassemblage de flux qui sont souvent utilisés dans les techniques d'évasion d'IDS. Ils fournissent à la sortie des données formatées de manière adéquate pour faciliter la détection avec la syntaxe des règles SNORT. L'engin de détection avec sa base de règles teste les paquets. Le résultat loggé peut être Packet I/O Totals, les statistiques de protocoles, d'actions, limites et verdicts dans son mode IDS. Le tableau 3.1 donne un exemple de règle et résume ses différentes options.

3.2.2 Stratégie de détection des IDS/IPS

La stratégie de détection d'un IDS est composée de deux grandes catégories : la détection de malveillances et la détection d'anomalies.

2. Berkeley Software Distribution : système OS dérivé d'UNIX et distribué par l'université de Californie, Berkeley

TABLEAU 3.1 – Les règles de SNORT

Ex. de règle SNORT	alert tcp any any -> 192.168.1.0/24 111 (content:"|00 01 86 a5|" ; msg:"mountd access" ;)	
Entêtes	action protocole IP src port src -> IP dest port dest	
	Actions	alert, log, pass, drop, reject, sdrop
	Protocoles	ip, icmp, tcp, udp
	Adresses IP	Format CIDR, liste d' @IP possible
	Ports	Numéro de ports standards
	Directions	-> ou <- ou <->
Options	(option 1:"argument" ; option 2:"argument" ; etc.)	
	Générales	msg, reference, classtype, etc.
	Payload	Content, rawbytes, depth, etc.
	Non-Payload	ttl, flags, flow, etc.
	Post-detection	logto, session, react, etc.

3.2.2.1 Détection de Malveillances (Misuse detection)

Dans les IDS qui utilisent la détection de malveillances, il y a une base de connaissances qui fournit au système des informations (règles) sur ce qui est malveillant. Les informations fournies contiennent des signatures d'attaques. Une signature, à l'image d'un *exploit*, représente un ensemble de caractéristiques identifiant une intrusion. Le système analyse alors constamment tout le réseau à la recherche d'une quelconque malveillance. Les mécanismes de détection de malveillances sont mis en œuvre par :

— **La recherche de motifs ou pattern matching**
La recherche de motif consiste simplement à rechercher des séquences alphanumériques caractérisant des attaques. Dans le cas d'une session, une opération de corrélation sur l'ensemble des paquets de la session est faite au lieu de faire la recherche paquet par paquet (statefull pattern matching). Cette méthode est basée sur des règles comme celui dans le tableau 3.1 contenant des caractéristiques de vulnérabilités publiées dans les CVE [10]. Étant donné que les attaquants masquent bien leurs attaques, créent des variations ou exploitent des vulnérabilités encore inconnues, l'efficacité n'est assurée que pour ce qui a été déjà défini dans la base de connaissances. Cette base doit être par conséquent constamment à jour.

— **L'analyse protocolaire**
Cette méthode est utilisée pour vérifier une utilisation des protocoles en conformité avec les RFC [3] . Cette méthode examine les champs et paramètres des paquets/entêtes et vérifie une utilisation suspecte en fonction des protocoles. Nous pouvons donner en exemple une taille inhabituelle d'entêtes ou de paquets, un formatage suspect de trames Ethernet, etc.

3. Request For Comment : Releases pour donner des spécifications de protocoles, des normes, etc.

3.2.2.2 Détection d'anomalies

Contrairement à la détection de malveillantes, les IDS basés sur la détection d'anomalies ont une connaissance de ce qui est normal grâce à une base de connaissances de profils. Ils recherchent et signalent toute activité ne respectant pas les caractéristiques des profils. Cette méthode est aussi appelée détection par analyse comportementale et dépend du comportement passé et nécessite un apprentissage. Cette analyse peut être mise en œuvre par des techniques statistiques/probabilistes en définissant :

— **Des seuils**
Ces seuils (indépendants des utilisateurs) sont fixés sur la base de ce qui est normal. Ce type de système décompte l'occurrence d'un évènement spécifique dans un intervalle de temps et déclenche une alarme si la valeur seuil fixée est atteinte. La durée maximale d'une session, le nombre de tentatives de connexion, le nombre de paquets par seconde, etc. peuvent avoir des seuils fixés. Cependant, il est très difficile de fixer un seuil pour un ensemble de personnes qui peuvent avoir des habitudes hétérogènes. Dans des contextes de ce genre, la détection par seuil est inefficace.

— **Des profils d'utilisateurs, de groupes ou d'utilisation de ressources**
La détection d'anomalies basée sur les profils utilise les caractéristiques du comportement passé des utilisateurs, analyse les données d'audits historiques suivant plusieurs métriques afin de déterminer le profil d'activité d'utilisateur moyen. Les métriques utilisées peuvent être des accès, des compteurs (temps, tentatives, etc.), l'utilisation des ressources (CPU, bande passante, etc.), etc. Les déviations sont alors détectées par l'analyse des données d'audit actuelles (fraîchement sauvegardés). Les données d'audits peuvent être :

- Les données d'audit déjà disponibles dans le système, mais souvent dans un format inapproprié, comportant les informations système, utilisateurs
- Les données d'audit créées spécifiquement pour la détection d'intrusion telles que les bases de données SIEM qui collectent les informations pertinentes.

Cette approche statistique peut être exécutée par les réseaux probabilistes bayésiens, par des modèles statistiques avec des réseaux neuronaux ou avec des systèmes experts combinés au Data-Mining [4] . Le test statistique peut se faire en calculant la moyenne et l'écart type pour déterminer un seuil, en utilisant des méthodes d'analyses multivariées, en utilisant le processus de Markov, etc. [2]. Les profils d'utilisateurs peuvent être combinés aux seuils pour une amélioration de la détection d'anomalies. Les profils d'utilisateurs sont vulnérables aux utilisateurs irréguliers ou dynamiques. Ils peuvent aussi être lentement habitués à un comportement intrusif. Les profils de groupes sont alors créés en étant moins vulnérables à l'irrégularité des utilisateurs ou encore à l'habituation lente à un

4. Domaine impliqué dans la détection d'intrusion permettant d'extraire des modèles descriptifs des énormes volumes de données d'audit [25]

comportement intrusif. En dehors de ces profils cités précédemment, nous avons aussi la possibilité de définir des profils d'utilisations de ressources ou encore d'exécution de programmes pour contrer les activités des malwares. Tous ces profils peuvent être statiques ou adaptatifs.

L'analyse de comportement peut être effectuée en utilisant des profils adaptatifs à base de règles (rule-based anomaly detection). Avec cette approche semblable à la détection statistique d'anomalie, l'IDS observe les évènements dans le système et définit un ensemble de règles pour définir le caractère suspect d'une activité. Il analyse les données historiques d'audit pour identifiant les profils et génère de manière automatique pour ces derniers les règles appropriées. Ensuite, l'IDS observe l'état ou le comportement actuel du système et fait la correspondance avec les règles pour vérifier la conformité. Cette approche ne nécessite pas de connaissance préalable de failles de sécurité.

L'analyse comportementale peut aussi être effectuée en utilisant une approche immunologique qui émule le système immunitaire humain dans ses propriétés de fonctionnement distribué et autonome. Le système expert recherche constamment le caractère malveillant dont il détient la connaissance tandis que le détecteur d'anomalies recherche le caractère normal dont il détient la connaissance. Les systèmes immunitaires artificiels quant à eux détiennent une connaissance de ce qui normal comme les détecteurs d'anomalie et recherchent constamment les activités malveillantes comme les systèmes experts [25].

3.2.2.3 Système Hybride

Chaque modèle a ses avantages et ses inconvénients. Il est donc naturel de vouloir combiner les modèles pour pallier les inconvénients des uns et des autres. La plupart des études récentes proposent des frameworks avec des combinaisons et obtiennent d'excellents résultats. Un exemple de framework mettrait la détection d'anomalies en amont pour séparer le trafic normal des anomalies pour ensuite appliquer une détection de malveillances sur les résultats de la première couche et vice-versa comme dans les études recensées dans [2].

3.2.2.4 Temps de détection et type de classification

Comme nous l'avons vu plus haut dans le fonctionnement du NIDS, ce dernier peut principalement fonctionner en mode TAP ou en mode inline. Le mode TAP correspond à une détection passive c.-à-d. post-infection et le mode inline correspond souvent à un détection en temps réel.

Dans les rapports de détection, l'IDS fournit des détails sur le type de trafic. Généralement, l'IDS sépare le trafic dans les rapports en trafic d'attaque et en trafic normal comme dans le cas de la détection binaire et de la détection d'anomalies. Maintenant,

certains des nouveaux IDS fournissent plus de détails avec plusieurs classes de trafic. La recherche va plus loin en incluant une classe pour le trafic inconnu.

3.2.3 Stratégie de déploiement des IDS/IPS

3.2.3.1 Architecture du déploiement

Pour les petits réseaux d'entreprise, il est fréquent de voir le déploiement d'un seul et unique IDS (mode standalone). Ce qui est suffisant pour les besoins de ce réseau. Cependant, pour les grandes entreprises avec plusieurs branches, déployer un unique IDS même pour chaque branche est insuffisant. Une approche collaborative a été proposée avec le déploiement de plusieurs IDS qui collaborent entre eux pour une vue globale du réseau de l'entreprise et une gestion centralisée. Cette approche collaborative peut être sous une architecture centralisée, sous une architecture hiérarchique ou encore sous une architecture distribuée. Dans chacune de ces architectures, nous pouvons avoir des modules d'agents distribués partout dans le réseau et sur des hôtes, donc des points de collectes et d'analyses multiples, et un ou des nœuds de gestion pour le traitement, la corrélation des évènements et la détection d'intrusion.

— **Architecture centralisée**
Dans cette architecture, nous avons un nœud de gestion centrale, comme dans le cas des topologies en étoile, qui reçoit les données collectées par les NIDS/HIDS dans le réseau. Cette architecture a un très bon taux de détection, mais présente l'inconvénient d'avoir un nœud central comme un point à défaillance unique Single Point of Failure (SPOF) qui désactive tout le système de détection en cas de panne. Et même s'il n'y a pas de panne, le nœud central peut être submergé d'alertes dans le cas d'une attaque massive. Il aura tendance à ignorer ou à délayer certaines alertes.

— **Architecture hiérarchique**
Pour corriger les inconvénients de l'architecture centralisée, la hiérarchie atténue l'effet qu'une panne du nœud central pourrait avoir en déployant des mini-nœuds centraux dans chaque branche par exemple. Ce mini-nœud aura les mêmes fonctions de traitement, de corrélation. Cependant, si ce mini-nœud tombe en panne, le système de détection de toute sa branche est désactivé.

— **Architecture distribuée**
Dans cette architecture, nous n'avons que des nœuds capables de collecter, d'analyser et de corréler les informations pour détecter les attaques locales. Ils ont aussi la capacité d'analyser et de corréler les informations provenant des nœuds voisins pour détecter les attaques globales.

3.2.3.2 Plateforme de déploiement

Comme nous l'avons vu dans la section des types d'IDS, une détection d'intrusion peut être un équipement réseau, une application dans un hôte, une application dans le Cloud (comme pour les XDR), dans les points d'accès (WIDS/WIPS) et les appareils mobiles. Les appareils IoT quant à eux n'ont pas assez de puissances pour analyser et traiter les informations, mais peuvent servir de points de collectes [28].

3.2.4 Stratégie d'évaluation des IDS/IPS

3.2.4.1 Métriques d'évaluation

Nous avons besoin de mesures pour quantifier la performance de détection des IDS. En matière de détection, on peut compter le nombre de vrais positifs (True Positive (TP)), le nombre de vrais négatifs (True Negative (TN)), le nombre de faux positifs (FP) et le nombre de faux négatifs (FN). Les vrais positifs représentent ce qui est détecté comme anormale sans erreurs et les vrais négatifs représentent ce qui est détecté comme normal sans erreurs. Tandis que les faux positifs représentent ce qui est détecté par erreur comme anormal et les faux négatifs représentent ce qui est détecté par erreur comme normal. Il y a un outil appelé matrice de confusion qui permet de visualiser ces métriques. Dans les matrices de confusion (figure 4.8), le nombre de faux positifs d'une classe est la somme de la ligne du trafic normal dans les vraies classes à l'exception de la valeur diagonale et le nombre de faux négatifs d'une classe est la somme de la colonne du trafic normal dans les classes prédites à l'exception de la valeur de la diagonale. Cette dernière est le nombre de vrais positifs. Le nombre de vrais négatives d'une classe est la somme de toute la matrice soustraite des vrais positifs, des faux positives et des faux négatives.

Il existe d'autres métriques composées telles que l'efficacité ou Accuracy (équation 3.1), la Precision (équation 3.3), la sensibilité ou Recall (équation 3.4), le taux de faux négatifs ou FNR (équation 3.6) et le taux de faux positives (équation 3.5). L'efficacité est le nombre d'attaques correctement prédit par rapport au nombre total d'attaques. La précision est le rapport entre le nombre d'attaques correctement classées et le nombre total d'attaques classées. La sensibilité mesure la capacité de notre système à identifier correctement les attaques. La métrique F1-Score est une moyenne harmonique de la sensibilité et de la précision.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.1}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3.2}$$

où

$$Precision = \frac{TP}{TP + FP} \tag{3.3}$$

et

$$Recall = \frac{TP}{TP + FN} \tag{3.4}$$

$$FPR = \frac{FP}{FP + TN} \tag{3.5}$$

$$FNR = \frac{FN}{TP + FN} \tag{3.6}$$

Certaines études préfèrent utiliser des courbes pour illustrer les résultats. Nous pouvons donner l'exemple de la Courbe Receiver Operating Characteristic (ROC). La courbe ROC peut avoir les FP ou le FPR ou encore la sensibilité sur l'axe des x et les TP, le TPR ou encore la précision sur l'axe des y. Dans la courbe ROC, les TP sont représentés en fonction des FP pour différents seuils. Chaque point sur la courbe ROC représente un couple FPR et TPR correspondant à un certain seuil de décision. Lorsque le seuil de classification varie, un point différent sur le ROC est sélectionné avec un FPR différent et un TPR différent. Un test avec une classification parfaite (pas de recouvrement des deux distributions) a une courbe ROC qui passe par le coin supérieur gauche (100 % de sensibilité, 100 % de précision). Dans le cas d'une détection d'anomalies, des seuils sont utilisés en combinaison avec des profils.

3.2.4.2 Consommation de ressources

La consommation en CPU, en GPU, en mémoire et en bande passante sont aussi des métriques d'évaluation. Le juste milieu entre un équipement performant et un équipement énergivore est essentiel surtout avec l'évolution vers les IDS basés sur l'IA qui vont requérir une grande puissance pour un fonctionnement optimale.

3.3 Challenge : la détection de l'inconnu

Malgré toute cette technologie pour stopper les attaques, celles zero-day restent un véritable problème. Les cybercriminels disposent de beaucoup d'outils et de techniques sophistiqués pour découvrir les vulnérabilités zero-day et préparer des exploits à utiliser ou à vendre sur le dark web.

3.3.1 SIEM, Automatisation et TI

3.3.1.1 SIEM

Le Security Information and Event Management (SIEM) est une solution de sécurité qui aide les organisations à prévenir tout incident de sécurité en découvrant les potentielles menaces et vulnérabilités. Les SIEM effectuent une gestion des journaux d'évènement en collectant, en analysant et en sauvegardant dans une base de données centralisée les logs et flux de données provenant des actifs, des applications, du Cloud, des utilisateurs, et des réseaux. Le SIEM intègre des fonctions d'agrégation de logs et de corrélation d'alertes permettant de déceler les vulnérabilités, localiser et atténuer les menaces à la sécurité de l'organisation. Le principe est d'exploiter la similarité dans les alertes dans le but de réduire la quantité d'alertes, d'améliorer la qualité de traitement et de l'exploitation, et de faciliter la compréhension de ces alertes. La finalité est de réduire le temps de détection et de réponse. Au-delà de ses fonctions précitées, le SIEM surveille constamment la sécurité et déclenche des alertes et des réponses en cas d'attaques par malware et d'attaque APT à l'image des IDS/IPS. Un SIEM est l'outil principal d'un SOC et travaille en complémentarité avec les IDS/IPS.

3.3.1.2 Automatisation

Avec l'évolution de l'IA, beaucoup de techniques ont vu le jour permettant d'automatiser certaines tâches. Dans le domaine de la sécurité, la solution Security Orchestration, Automation and Response (SOAR) a été mise en œuvre pour automatiser les routines de sécurité et les activités fastidieuses et voraces en temps. Le principe d'orchestration est de connecter les outils ensemble pour éviter une gestion par outils et pour atténuer l'effet de fatigue causé par la grande quantité d'alertes. L'équipe NOC/SOC peut effectuer toutes les routines pour chacun des produits de sécurité dans le chapitre précédent directement sur l'interface SOAR. En plus, ces opérations peuvent être automatisées grâce à des organigrammes à exécution automatique de l'outil SOAR appelés playbooks. L'outil SOAR est d'ailleurs souvent intégré à la solution SIEM étudiée précédemment et permet de répondre automatiquement en cas d'incidents.

3.3.1.3 Threat intelligence

La Threat Intelligence (TI) est une veille de sécurité fournissant des renseignements sur les menaces et attaques et permettant de suivre leur évolution en continu dans le but de connaitre pour mieux répondre et surtout mieux anticiper. Une autre description définit la Threat Intelligence comme le processus de changement de statut d'une menace de "unknown unknown" à "known unknown" en découvrant leurs existences, puis de "known

unknown" à "known known", état dans lequel la menace est bien comprise et atténuée. Le "unknown unknown" est l'ensemble des menaces et risques dont nous n'avons pas idée de leur existence. Le "known unknown" est l'ensemble des menaces et risques dont l'existence est connue, mais sur lesquels nous n'avons pas une maîtrise, et enfin le "known known" qui est l'ensemble des menaces et risques connus et maîtrisés. La notion clé ici est le terme renseignement qui est un ensemble d'informations et de données corrélées, et vérifiées par des experts, permettant de reconstituer une suite d'évènements. Ces renseignements sont partagés entre bases d'intelligences et avec les organisations. Ces bases d'intelligences peuvent être open source comme les bases de données CVE, par exemple la base NVD de la NIST [29] ou encore TALOS [10]. Elles peuvent être aussi sous forme de services fournis par des fournisseurs ou des équipes d'experts comme MANDIANT [19] ou THE-MITREE [20]. Plus de détails sont donnés dans le livre blanc de Chismon and Ruks [30].

3.3.2 Honeypot/Honeynet et Sandboxing

3.3.2.1 Honeypot/Honeynet

Un honeypot est un point de leurre comme un hôte ou généralement un serveur avec un faible niveau de sécurité. Il est utilisé dans le but de leurrer les cybercriminels ou les insiders avec une cible plus attractifs que le réseau réel de l'organisation, mais aussi de récolter des informations sur les méthodes d'attaques notamment inconnues (zero-day exploits) utilisées par les attaquants. Un honeynet quant à lui est un réseau souvent virtuel ou une zone de honeypots virtuels au sein d'un réseau d'organisation.

3.3.2.2 Sandboxing

Étant donné que l'approche de la détection par signature traditionnelle n'est pas très efficace contre les malwares polymorphes, nous avons développé une autre approche qui nous permet de détecter une famille entière de malwares polymorphes en utilisant une seule signature. Cela est possible à cause des similarités au sein d'une famille. Cette approche est très efficace contre les malwares et variations connues. Pour gérer ceux encore inconnus, des produits appelés sandbox ont été développés. Ces produits placent tout fichier suspect dans un environnement isolé comme le sandbox de Windows qui est presque aussi pareil que l'OS. Le comportement est analysé dans cet environnement et si des actions malveillantes sont détectées, alors le fichier est signalé comme malware. Cette approche est appelée détection heuristique. Au-delà du Sandboxing, l'avenir de la détection de logiciels malveillants jusque-là inconnus inclut l'utilisation de l'IA et du ML pour évaluer le potentiel de menace de sécurité des fichiers en temps réel.

3.3.3 Vulnérabilités zero-day et Bug Bounty

3.3.3.1 Vulnérabilités zero-day

Récemment, plusieurs attaques de grande ampleur comme Stuxnet ont eu lieu et ont en plus exploité des vulnérabilités d'alors inconnues. Ces dernières ont suscité un grand intérêt et sont devenues une des préoccupations principales de la recherche en cybersécurité. La recherche des vulnérabilités devient de plus en plus importante et plusieurs frameworks pour rechercher des vulnérabilités zero-day ont été proposés. Selon [31], si des vulnérabilités sont publiquement divulguées avant la disponibilité des correctifs, le fournisseur exécute une évaluation de risque sur les exploits zero-day et les potentielles attaques contre son produit vulnérable. La solution optimale est de publier dès que possible un correctif pour toute vulnérabilité découverte. Mais le coût de développement des correctifs est le premier facteur de la lenteur de la publication des correctifs. Des évènements, comme celui d'Equifax [9], ont montré qu'une publication dès que disponible peut être fatale étant donné que certains administrateurs ne suivent pas constamment la publication des correctifs. Le deuxième facteur est alors le choix d'effectuer des publications périodiques. Les principaux problèmes avec ces releases périodiques sont la peine que les fournisseurs ont pour prioriser certaines vulnérabilités sur d'autres, mais aussi qu'entre temps, un attaquant peut trouver et exploiter une vulnérabilité zero-day.

3.3.3.2 Détection de vulnérabilité zero-day

Le cycle de vie d'une vulnérabilité zero-day s'étend de la découverte de son existence à la publication de son correctif. Entre-temps, il peut y avoir la création de l'*exploit*, le lancement d'attaques, des dommages dans l'organisation. C'est pourquoi, des auteurs comme Singh et al. [32] se lancent dans la découverte de vulnérabilités. Ces auteurs ont proposé un framework avec une architecture à trois couches qui détecte, analyse et priorise les vulnérabilités pour venir en aide aux fournisseurs.

1. Dans ce modèle proposé, le trafic réseau est initialement filtré par un SNORT IDPS qui bloque les attaques connues grâce à leur signature pour limiter la charge du réseau.
2. Ensuite, ils ont placé un honeypot «passif» pour logger le trafic qui a réussi à traverser la détection par signature. Ensuite, ce trafic est étiqueté et combiné aux informations collectées par le honeypot.
3. Cet ensemble est à nouveau soumis à une détection d'anomalies (un autre SNORT NIDS) dans le Detection Engine du framework. Il est comparé avec le trafic généré par les hôtes sûrs du banc d'essai pour déceler tout comportement suspicieux.
4. Si le trafic est classé comme normal, il est collecté pour mettre à jour la base de

données de SNORT NIDS. Dans le cas contraire, il est enregistré dans la base de données Zero-day dans la couche physique de leur architecture. Ce processus est le «zero-day attack path generator».

Quant au mécanisme de priorisation de vulnérabilités, il est basé sur une analyse de risques. Le «graphe» d'attaque obtenu précédemment est une chaîne d'exploits de vulnérabilités possibles. Ce qui permet de localiser les failles de sécurité du système. Les auteurs se basent sur une approche probabiliste notamment avec la vraisemblance pour déterminer le risque d'exploitabilité d'une vulnérabilité étant donné que l'impact est inconnu à cause du caractère zero-day. L'analyse de risque est basée sur le vecteur d'attaque, la complexité de l'attaque, entre autres pour prédire ou déterminer la fréquence ou la vraisemblance d'exploitation d'une vulnérabilité. Selon ce raisonnement, plus la vraisemblance est élevée, plus la correction est prioritaire. Avec les expériences menées en émulant un caractère zero-day sur des équipements non à jour, ils ont obtenu un pourcentage de détection d'attaques zero-day de 96 % avec un faible FAR de 0,3 %.

3.3.3.3 Bug Bounty

Les Bug Bounty sont des programmes qui regroupent des chercheurs en sécurité. Ces programmes sont des plateformes entre les chercheurs et les fournisseurs. Ils donnent des primes financières pour la découverte de vulnérabilités zero-day et communiquent après avec le fournisseur concerné pour partager les informations en privé. Une échéance est donnée aux fournisseurs pour développer des correctifs avant de rendre publiques ces vulnérabilités. À ce jour, les Bug Bounty constituent la source principale de découverte de vulnérabilités zero-day.

3.4 Limites des IDS et Stratégie de détection basée sur l'IA

3.4.1 Vulnérabilités et limites des IDS

3.4.1.1 Vulnérabilités des IDS

Malgré leur capacité à détecter les infiltrations, les IDS sont cependant vulnérables à certaines techniques dites d'évasion qu'utilisent les cybercriminels pour contourner la détection. Tout d'abord, il faut savoir que les attaquants ont la capacité de détecter les IDS/IPS en analysant la latence dans les réponses aux requêtes, les réponses aux attaques ou en recevant une réponse à une requête ICMP avec le champ d'adresse MAC à zéro. Dans le dernier cas, l'IDS répond à la requête s'il est en mode promiscuité. Pour contourner la

détection d'un IDS, les cybercriminels peuvent :

- Cacher du trafic malveillant dans les paquets en utilisant le mécanisme de fragmentation. Étant donné que chaque type d'équipement peut avoir son propre mécanisme de reconstitution de paquets, les attaquants peuvent adapter des mécanismes d'écrasement de fragments aux mécanismes de reconstitution d'un IDS qui va écraser les fragments malveillants avec des fragments normaux. Par contre, la cible dans sa reconstruction peut restituer les paquets tels que conçus par les attaquants. Ces derniers peuvent aussi insérer des fragments normaux avec un TTL adapté au réseau entre les fragments malveillants. Ces paquets reconstruits au niveau IDS ne vont matcher aucune signature. Cependant en traversant le réseau, les fragments normaux vont être supprimés à cause de leur TTL et la cible va recevoir et reconstruire les paquets d'attaques qui seraient normalement détectés s'ils étaient reçus comme tels par l'IDS. Les attaquants peuvent aussi exploiter la durée de vie des fragments dans la mémoire cache de l'IDS si elle est plus courte pour attaquer une cible ;
- Utiliser des paquets UDP ou ICMP pour lancer un déni de service par épuisement de ressources sur l'IDS. Ce dernier aura du mal à détecter les attaques enfuies dans l'énorme quantité de données ;
- Utiliser des techniques de dissimulation. La norme Unicode/UTF-8 permet de coder un caractère par plusieurs formats différents possibles. Les cybercriminels peuvent également utiliser des données à double codage, avec la norme et des opérateurs XOR pour obtenir un caractère polymorphe, augmentant ainsi le nombre de signatures nécessaires pour détecter l'attaque ;
- Utiliser le chiffrement pour exploiter l'incapacité des IDS à analyser le trafic crypté.

Il faut noter que ces techniques sont moins efficaces contre la détection d'anomalies et les IPS qui présentent des améliorations des IDS en plus de la capacité à bloquer les attaques.

3.4.1.2 Limites des IDS

Les IDS sont capables de détecter une grande variété d'attaques. Cependant, ils présentent plusieurs inconvénients selon le format et plusieurs limites selon la stratégie de détection utilisée. Les inconvénients et les limites sont résumés respectivement dans les tableaux 3.2 et 3.3.

Tableau 3.2 – Inconvénients des IDS

Type de IDS	NIDS	HIDS
Inconvénients	Équipement dédiée requis	Consomme les ressources de l'hôte
	Beaucoup de ressources et de CPU	Installation sur chaque hôte
	Débit d'analyse de flux limité	Analyse locale uniquement
	Incapacité d'analyse de flux crypté	Retards d'alertes

Tableau 3.3 – Limites des IDS

Stratégie de détection	
Basée sur la signature	Mises à jour fréquentes nécessaires.
	Incapacité de détecter des variations d'attaques.
	Incapacité de détecter les attaques zero-day.
	Pas de détails sur les attaques.
	Ne convient pas pour la détection d'attaques en plusieurs étapes.
Basée sur les profils	Taux de Fausses alarmes élevé.
	Profils normaux difficiles à définir.
	Alertes non classées - pas de détails sur les attaques.
	Entraînement délicat - temps d'entraînement long
	Énorme quantité de fausses alarmes en cas de changements.
Basée sur les règles	Génération de règles dépendant des données d'audit
	Exigence d'un grand nombre de règles pour déterminer toutes les attaques possibles.
	Utilisation élevée du CPU car les règles utilisent le pattern matching.

3.4.2 Stratégie de détection basée sur l'IA

Dans la section sur la stratégie de détection, nous avons omis volontairement la capacité de l'IA plus précisément du Machine Learning (ML) à implémenter les mécanismes de détection de malveillances et de détection d'anomalies. En effet, les modèles de ML plus spécialement les réseaux de neurones ont montré une très bonne capacité à détecter des malveillances ou des anomalies. Des surveys comme ceux de Ansam et al. [2] et de Lansky et al. [3] ont recensée un ensemble d'études avec les techniques de détection citées dans la section sur les stratégies de détection des IDS, mais aussi avec les méthodes basées sur le ML et sur le DL. Ces modèles peuvent implémenter des mécanismes de régression ou de classification. La régression est une technique d'analyse de données simples utilisée en Machine Learning pour séparer deux catégories de données. Elle peut être linéaire ou non linéaire. Elle est souvent utilisée dans la détection d'anomalies pour séparer les données normales des anomalies en définissant un seuil ou un plan de séparation. La classification quant à elle permet de classer les données de caractéristiques communes. Elle peut être

binaire ou multiclasse. Pour plus de détails sur les modèles de Machine Learning, il faut se référer à l'étude de Ansam et al. [2] car nous allons nous concentrer sur les modèles de DL. Les modèles de Deep Learning [4] peuvent être :

- Génératives (Auto-Encodeur, RBM [5] , DBN [6] ,RNN, etc.) avec un apprentissage non supervisé sans utilisation d'étiquettes pour reconnaître les classes ;
- Discriminatives (DNN, CNN) avec un apprentissage supervisé avec utilisation d'étiquettes pour reconnaître les classes ;
- Hybrides (GAN illustré dans la figure 1.3)

3.4.2.1 Deep Neural Network - DNN

Le Deep Neural Network (DNN) est une amélioration du réseau de neurones artificiels. Le nom vient du fait qu'il est composé de plusieurs couches de neurones. Ce réseau a la capacité de prendre en entrée des données multidimensionnelles à grande échelle, d'en apprendre les caractéristiques pertinentes et de déceler les corrélations entre elles. Ce type de réseau possède la capacité de trouver le modèle adéquate pour les données en entrée avec les coefficients associés. Il possède aussi la capacité d'évaluer son apprentissage afin de trouver les coefficients optimaux à chaque itération et de se mettre à jour grâce au mécanisme de Back-Propagation. Il peut être utilisé pour des tâches de classification ou des tâches d'extraction de caractéristiques. Leur implémentation peut se faire grâce à des frameworks comme TensorFlow, PyTorch, etc.

3.4.2.2 Convolutional Neural Network - CNN

Le réseau CNN est un cas particulier du DNN dans l'extraction de caractéristiques. On peut dire que les CNN sont des réseaux à plusieurs couches avec en amont des couches qui effectuent des opérations de convolutions et de filtrage pour effectuer l'extraction de caractéristiques et à la fin des couches semblables au réseau DNN pour effectuer la classification. Ce réseau a été développé pour le traitement et la classification d'images, mais présente d'excellents résultats dans la classification d'attaques.

3.4.2.3 Recurrent Neural Network - RNN

Le Recurrent Neural Network (RNN) est un autre cas particulier du réseau DNN composée des neurones dotés de mémoire pour prendre en compte les états précédents du réseau dans son apprentissage. Il a été développé pour s'adapter aux séries temporelles

5. Restricted Boltzmann Machine
6. Deep Belief Network

dans lesquelles les données ont une corrélation dans le temps. Les RNN sont particulièrement efficaces sur les données réseau qui ont un caractère temporel. Ils existent des améliorations du RNN comme le réseau Long Shot Time Memory (LSTM) dans lequel les neurones ont une mémoire plus grande pouvant prendre en compte beaucoup plus d'états précédents. Il y a aussi le Gated Recurrent Unit (GRU) qui est une version simplifiée de LSTM.

3.4.2.4 Auto-Encodeur - AE

L'AE est un réseau composé de deux DNN particuliers séparés par un nœud d'étranglement appelé espace latent ou Bottleneck. Le processus d'apprentissage et la structure de l'AE sont expliqués plus en détail dans le chapitre suivant. Entre-temps, il faut noter qu'il existe plusieurs variations d'AE comme les Deep ou Stacked AE (DAE) qui ont plus de couches que l'AE, les Sparse AE qui ont des contraintes de parcimonie ainsi que les Denoising AE qui sont utilisés pour nettoyer les impuretés dans les données d'entrée. Les contraintes de parcimonie visent à produire une sortie avec une moyenne faible en désactivant un grand nombre de neurones la plupart du temps.

Pour plus de détails sur les autres modèles de DL existants, se référer aux études de Aldweesh et al. [4] et de Lansky et al. [3] qui enquêtent sur les modèles de DL disponibles dans la littérature.

Ces techniques basées sur l'IA peuvent être utilisées pour l'extraction des caractéristiques pertinentes pour faciliter le travail du module d'analyse et de détection. Elles peuvent aussi être utilisées à la fois pour l'extraction des caractéristiques pertinentes et pour l'analyse et la détection. Les modèles développés dans nos expériences (chapitre suivant) effectuent les deux processus et peuvent remplacer le module d'analyse et de détection. Cette amélioration aura pour effet :

- D'augmenter le débit d'analyse du module d'analyse et de détection ;
- De donner à l'IDS la capacité de prendre en charge une énorme quantité de données ;
- D'améliorer la sensibilité, la précision de l'IDS ;
- De diminuer le taux de faux négatifs et le taux de faux positifs ;
- De diminuer le temps de détection ;
- De donner à l'IDS la capacité de détecter les attaques zero-day ;
- De donner des détails sur les attaques à l'administrateur avec une classification des attaques.

La figure 3.6 offre une vue complète des techniques utilisées dans la stratégie de détection des IDS.

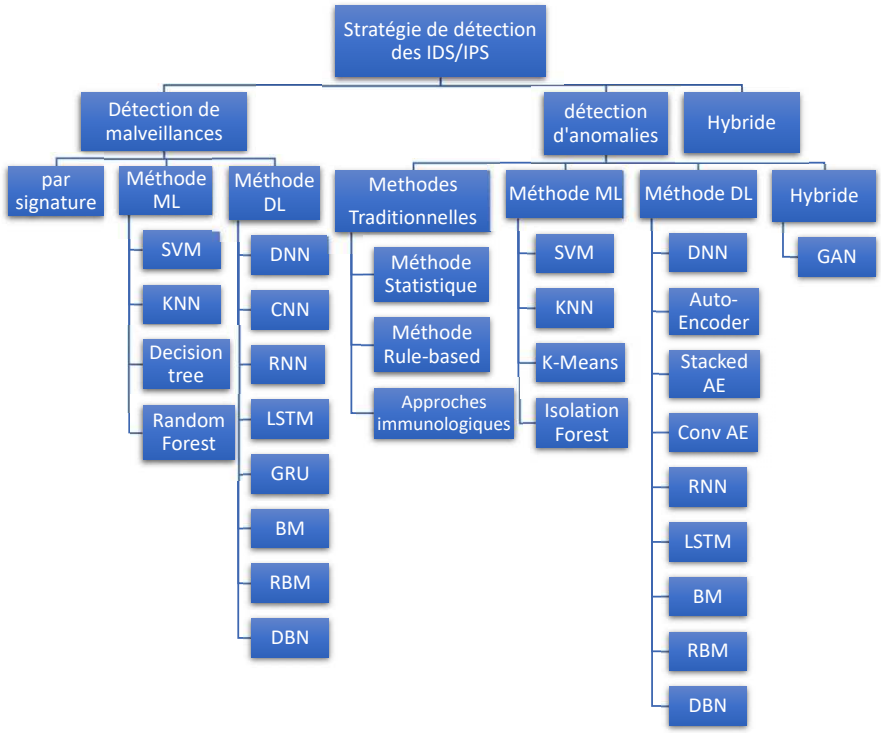


Figure. 3.6 : La Stratégie de détection des IDS

3.5 Conclusion

Dans cette partie de notre travail, nous avons étudié une grande variété de contre-mesures, détectives, correctives. L'IDS/IPS qui constitue notre objet d'étude est présente dans nos réseaux sous plusieurs formes. Il effectue une suivie constante du réseau ou de l'hôte où il est déployé afin de déceler toute intrusion ou attaque. A l'occurrence d'un évènement, l'IDS déclenche tout un ensemble de mécanismes afin de permettre à l'administrateur ou à l'opérateur NOC/SOC d'appliquer les contre-mesures. La stratégie de détection utilisée par l'IDS est importante à comprendre, car elle détermine ses capacités de détection, le temps de détection qui est un paramètre important et ses limites. Nous avons vu que l'IDS basée sur les signatures est très efficace contre les attaques connues tandis que l'IDS basée sur la détection d'anomalies présente en plus la capacité de détecter les attaques jusqu'alors inconnues. Cependant, la stratégie de détection des IDS/IPS nécessite encore des améliorations dans la capacité de détection des attaques connues et inconnues, dans la réduction des erreurs de détection, dans la diminution du temps de détection, dans l'augmentation du débit d'analyse. La recherche s'est orientée vers l'IA à travers le Deep Learning pour repousser les limites de l'IDS et dans la mesure du possible, obtenir d'autres fonctionnalités. Dans ce qui suit, nous proposons de remplacer le moteur d'analyse et de détection de l'IDS par un modèle de DL pour tenter de repousser encore les limites de l'équipement.

Chapitre 4

Stratégie de détection de l'inconnu par Deep Learning

4.1 Introduction

Nous avons vu que les IDS actuels détectent efficacement la plupart des attaques connues. Les challenges de ces équipements résident dans la capacité de détecter les attaques inconnues (zero-day), mais aussi de déclencher moins de fausses alarmes. En plus, à l'ère du big data, un très gros volume de données de tout type traverse nos réseaux. C'est pourquoi les IDS doivent aussi être capables de gérer cette quantité de données pour détecter les attaques et y répondre à temps. Dans le challenge de détection de l'inconnue, des solutions comme le SIEM, et la Threat Intelligence ont été adoptées, mais celles-ci présentent des inconvénients de coûts et d'expertises.

En pratique, malgré leur utilité, les IDS déclenchent trop de fausses alarmes et sont incapables de détecter certaines attaques. Ajouté à ces problèmes, le paysage des cybermenaces ne cesse de s'étendre et les cyberattaques deviennent de plus en plus intelligentes, rapides et furtives. La recherche commence alors à tester des techniques d'IA et le DL reste comme le candidat idéal pour :

- Gérer la haute dimensionnalité et la quantité de données ;
- Apprendre et extraire lui-même les caractéristiques du trafic réseau et leurs corrélations ;
- Prédire les attaques en conséquence selon les labels ou le seuil et fournir des rapports de détection très détaillés.

Dans notre étude, nous proposons une détection d'anomalies non supervisée par un Auto-Encodeur profond (DAE) et une classification multiclasse de trafic par un réseau de neurones convolutionnels (CNN). Nous utilisons comme données d'expériences, l'ensemble de données CSE-CIC-IDS2018 qui est très proche des données de réseaux actuels avec une

énorme quantité d'échantillons (16 232 943, dont 82 % de trafic normal et 14 catégories d'attaques récentes). Le reste de ce chapitre est organisé comme suit. Dans la deuxième partie, nous étudions l'IDS basé sur DL, en particulier les études basées sur CNN et DAE avec CSE-CIC-IDS2018. Dans la troisième partie, nous donnons des détails sur nos méthodes proposées. Dans la quatrième partie, l'analyse exploratoire des données, le pré-traitement pour chaque modèle et les résultats de nos expériences sont détaillés. Nous discutons nos résultats dans la cinquième partie. La sixième partie conclut ce chapitre.

4.2 Étude de performance des différents modèles IDS basés sur le CNN et l'auto-encodeur pour la détection d'intrusion

Nous avons trouvé dans la littérature de nombreuses études récentes sur la détection d'intrusion basée sur le Deep Learning, la détection zero-day, la détection d'anomalies, la classification multiclasses, etc. Pour poser des limites à notre étude, nous décrivons en priorité certaines études qui ont utilisé le jeu de données le plus récent (CSE-CIC-IDS2018).

Dans l'article Kim et al. [33], les auteurs ont proposé un modèle basé sur le CNN pour la détection d'intrusion. Ils ont formé un CNN à 2 couches avec 79 caractéristiques de données de l'ensemble de données CIC-IDS2018 subdivisé en 10 sous-ensembles de données avec une subdivision par jour. Elles ont formé et validé leur modèle avec chaque sous-ensemble avec les données mises au format image. Avant le pré-traitement, leur modèle fonctionne bien dans dans la classification des trafics bénigne, DoS Hulk, DoS Slow HTTPtest, botnet, toutes les attaques DDoS et brute force (SSH et FTP). Cependant, le modèle présente des performances très médiocres sur les attaques Web, l'infiltration, le DoS GoldenEye et le DoS Slowloris. Leur modèle peine clairement contre le déséquilibre des classes. Elles ont également présenté des résultats avec quelques améliorations après pré-traitement.

Dans leur article, Yong and Bo [34] ont proposé un modèle CNN basé sur le modèle Inception de l'équipe Google. Le CNN à 9 couches est évalué avec le jeu de données KDDCUP99 avec des données encodées et mises à l'échelle. Ils obtiennent respectivement 94,11 %, 2,18 % et 93,21 % d'efficacité, de taux de FP et de taux de détection. Leur CNN doit être évalué sur des ensembles de données plus à jour. Le modèle doit également être amélioré comme les auteurs l'ont dit en termes de taux de FP.

Khan et al. [35] ont entraîne un 2D-CNN à 3 couches avec un jeu de données KDD99. Ils ont utilisé peu de données avec seulement 10 % de l'ensemble de données et l'efficacité comme seule mesure. Le modèle a atteint 99,23 % d'efficacité, mais après 800 itérations.

Les auteurs n'ont pas donné des détails sur les faux positifs et les faux négatifs.

Azizjon et al. [36] ont entraîné un 1D-CNN de 1 à 3 couches avec le jeu de données UNSW-NB15 dans leurs études. Ils ont également combiné un réseau 1D-CNN et un réseau LSTM. Ils ont choisi 43 caractéristiques de cet ensemble de données déséquilibré pour alimenter leurs modèles. Ils ont obtenu des efficacités de 76,76 %, 77,99 %, 80,33 % et 79,57 % pour le 1D-CNN (1 à 3 couches) et 1D-CNN+LSTM respectivement et des pourcentages de Recall exceptionnels entre 98 et 99,61 % pour leurs différentes configurations. On note qu'il y a un taux élevé de faux positifs et un très faible taux de faux négatif. Les auteurs ont ensuite utilisé une technique de suréchantillonnage aléatoire pour équilibrer les classes et ont refait l'expérience avec les mêmes modèles. Les expériences avec des classes équilibrées montrent des améliorations en termes de précision (environ 87,53 % pour leur meilleur modèle) réduisant le taux de fausses alarmes. Mais le taux de faux négatifs augmente, car le pourcentage de Recall toujours très bon a diminué d'environ 2 %. Cette étude illustre bien le problème du déséquilibre des données pour le deep learning.

Dans leur article, Kim et al. [37] ont proposé un IDS réseau (NIDS) basé sur un CNN contre les attaques DoS. Pour leur étude, ils ont utilisé l'ensemble de données KDD & CSE-CIC-IDS2018 converti au format image en couleurs RVB et image en niveaux de gris pour entraîner 18 configurations différentes de CNN. Ils obtiennent des résultats intéressants avec une efficacité élevée d'environ 99 % avec les deux ensembles de données. Cependant, leur processus d'évaluation n'est pas homogène, car ils n'ont pas utilisé les mêmes métriques pour la classification binaire (F1-SCORE) et la classification multi-classes (l'efficacité seulement qui ne suffit pas). Ils n'ont donné, pas non plus, des résultats détaillés des expériences avec CSE-CIC-IDS2018.

Yoo et al. [38] ont proposé le même modèle CNN que les auteurs précédents, mais avec une méthode de pré-traitement différente, car ils ont utilise une méthode de discrétisation de caractéristiques avant de convertir les données en images. Ils ont comparé leur proposition avec un 1D-CNN et ont montré que la discrétisation réduit le temps d'apprentissage des modèles. Les auteurs ont utilisé les ensembles de données NSL-KDD et CSE-CIC-IDS2018, et leurs résultats présentés sur l'ensemble de données CSE-CIC-IDS2018 en pourcentage moyen de 98 % (Accuracy, Precision, Recall) lorsqu'ils ont utilisé la discrétisation uniquement sur les caractéristiques continues et environ 97 % lorsqu'ils ont utilisé la discrétisation à la fois sur les caractéristiques continues et celles qui sont discrètes. Cependant, leur framework a montré de mauvaises performances sur l'ensemble de données NSL-KDD.

Les auteurs Kang et al. [39] ont proposé un modèle de détection d'intrusion basé sur l'auto-encodeur et le modèle XGBoost. Leur structure de modèle a un module de pré-traitement avec nettoyage des données, suréchantillonnage, sélection de caractéristiques par le modèle Random Forest et regroupement de caractéristiques, un module

d'auto-encodage entraîné avec des données normales et testé avec des données normales et anormales. Le seuil et l'erreur de reconstruction obtenus sont utilisés par le classificateur XGBoost pour la détection. Leur modèle proposé présente de très bonnes performances, notamment le Recall. Cependant, le modèle a du mal à détecter les attaques par infiltration et les attaques par injection SQL.

Catillo et al. [40] ont proposé le ZED-IDS et son amélioration 2L-ZED-IDS dans l'article [41] basé sur un stacked/deep auto-encodeur (SAE). Le ZED-IDS est un SAE unique qui présentent des bonnes performances sur les deux ensembles de données les plus récents, mais déclenche de nombreuses fausses alarmes (4,9 % comme le taux le plus élevé d'attaque DoS). Afin de diminuer le taux de fausses alarmes (FAR), les auteurs ont proposé un 2L-ZED-IDS à 2 couches. Dans leur dernier modèle, ils utilisent la sortie du premier SAE dans la première couche pour entraîner et évaluer deux autres SAE dans une deuxième couche. L'un est entraîné avec le trafic normal et les faux négatifs, le second est entraîné avec les trafics classés attaques et les faux positifs provenant de la sortie de la première couche. Leurs résultats montrent que le FAR a diminué de plus de 50 %. En effet, les autres mesures de performances ont également légèrement augmenté.

4.3 Approches méthodologiques

Dans cette section, nous décrivons les approches utilisées pour le pré-traitement des données afin d'entraîner et d'évaluer nos modèles proposés. Nous avons commencé par une analyse exploratoire des données (EDA) qui nous aide sur le pré-traitement des données. À la fin, les données sont mises sous un format approprié pour entraîner et évaluer respectivement chaque modèle.

4.3.1 EDA et pré-traitement

Une EDA est un début et une étape importante pour des études comme la nôtre. Une EDA sur CSE-CIC-IDS2018 nous permet de mieux comprendre les données, afin de savoir comment bien les exploiter. Cette étape est souvent négligée, mais joue un rôle essentiel dans la conduite d'un bon pré-traitement des données. Notre EDA est passée par l'analyse de la forme, l'analyse du contenu, l'analyse de la distribution du trafic, l'analyse des caractéristiques et la corrélation.

Il a été montré que dans le domaine de l'IA, le pré-traitement impacte grandement sur les performances d'un modèle. Les auteurs Kim et al. [33] ont bien illustré ses impacts sur les performances de leur modèle CNN. Avant d'alimenter les modèles avec des données, notre pré-traitement est passé par la recherche de Not a Number (NaN), les valeurs dupliquées et infinies à éliminer ou à imputer, l'encodage des étiquettes, la séparation

des données par étiquette et le suréchantillonnage aléatoire pour le cas de la classification des attaques, le découpage de l'ensemble de données en ensemble pour l'entraînement, en ensemble pour la validation et en ensemble pour le test, éliminant les caractéristiques inutiles et normalisant les données. Avant l'EDA et la description du pré-traitement, plus de détails sur les modèles proposés sont donnés.

4.3.2 Modèles

Nous proposons une étude des performances de classification du trafic d'un CNN 1D pour le jeu de données CSE-CIC-IDS2018. En parallèle, nous proposons une étude de détection d'anomalies pour le même jeu de données par un DAE.

4.3.2.1 La détection d'anomalie avec l'auto-encodeur

Pour la détection d'anomalies, nous utilisons un DAE dans un apprentissage non supervisé. Dans la phase d'encodage, en réduisant la dimensionnalité des données d'entrée, le DAE apprend les caractéristiques les plus pertinentes des données d'entraînement au niveau de l'espace latente ou bottleneck. Dans la phase de décodage, avec la représentation de cette couche, le DAE essaie de reconstruire les données tout en minimisant les pertes ou l'erreur de reconstruction (RE) à chaque itération. Le RE est calculé avec l'erreur quadratique moyenne (Mean Squared Error (MSE)) ou l'erreur absolue moyenne (Mean Absolute Error (MAE)) entre les données d'entrée X_i et les données de sortie $\hat{X}_i$. En résumé, un DAE compresse les données et construit sa propre représentation latente, puis reconstruit les données dans leur représentation d'origine en essayant de minimiser le RE en ajustant automatiquement les coefficients de neurones a_i. Le DAE définit alors un seuil entre données normales et anormales. Le seuil est calculé avec l'équation (4.1).

$$threshold = \mu(\mathcal{L}) + n\sigma(\mathcal{L}) \text{ or } threshold = \max(\mathcal{L}) \tag{4.1}$$

où μ est la moyenne des pertes d'entraînement ($\mathcal{L}$), σ l'écart type de ($\mathcal{L}$) et $n = 1, 2$ ou 3.

Notre DAE dispose d'un encodeur à 5 couches cachées, d'une espace latente ou bottleneck et d'un décodeur à 5 couches cachées. Nous avons inséré deux couches de batch normalisation entre les couches dans l'encodeur et deux dans le décodeur pour accélérer la convergence et pour éviter l'overfitting. Les dimensions d'entrée et de sortie sont égales au nombre de caractéristiques (figure 4.1).Nous avons utilisé Adam comme fonction pour l'optimisation et MSE comme fonction de perte pour la détection des anomalies par le DAE.

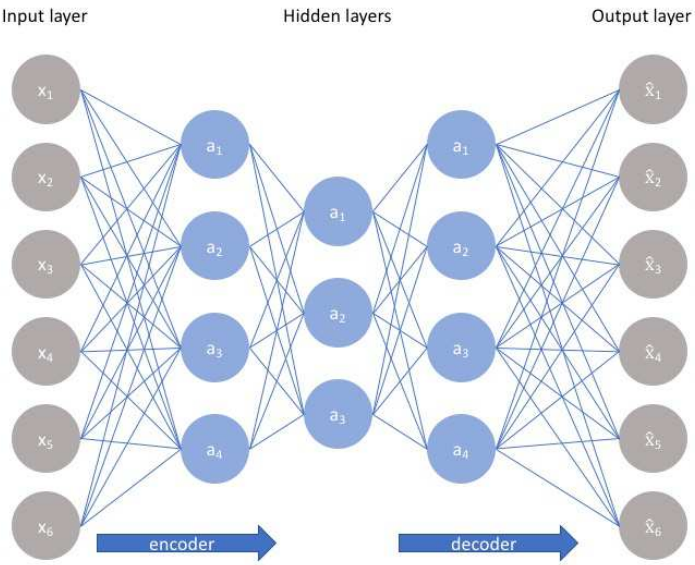


Figure. 4.1 : Architecture du Deep Autoencoder [42]

4.3.2.2 La classification avec le réseau CNN

Pour la classification des attaques, nous proposons un CNN avec une pile à trois niveaux de couches Conv1D et MaxPooling1D. Chaque niveau de la pile a :

- Une couche Conv1D avec 64 filtres, 6 comme valeur du kernel size et un format de (N°. de features, 1) comme format d'entrée. La couche de convolution extrait les caractéristiques les plus complexes des modèles à partir des données ;
- Une couche batch normalisation qui aide le réseau à converger plus rapidement vers le minimum global ;
- Une couche MaxPooling1D, avec une taille de pool de 3 et 2 strides, qui sélectionne la valeur maximale de la région où le filtre est appliqué. La sortie de conv1D est compressée à une taille plus petite par la couche de MaxPooling.

Il vient ensuite la couche Flatten qui remodèle les données de plusieurs dimensions à une dimension, suivie de couches entièrement connectées (Fully connected) avec 64 unités et le nombre d'étiquettes égale au nombre d'unités pour la couche de sortie (figure 4.2). La classification est faite par les couches entièrement connectées. Nous avons utilisé la fonction d'activation SoftMax dans la couche de sortie et RELU dans toutes les autres couches. La couche de Batch Normalization dans chaque couche convolutionnal est également incluse pour éviter le surajustement ou overfitting. Nous avons la possibilité d'utiliser Conv1D et MaxPooling1D pour entraîner le modèle avec des données unidimensionnelles ; par conséquent, nous avons choisi de ne pas ajouter d'étape supplémentaire de pré-traitement qui remodèle les données au format image et augmente le temps de traitement. Nous avons utilisé Adam comme fonction pour l'optimisation, Categorical Cross Entropy comme fonction de perte pour la classification des attaques.

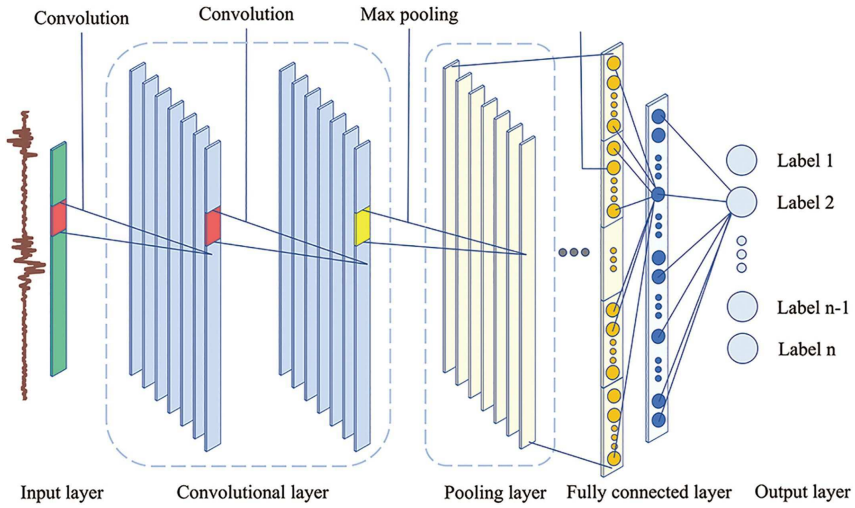


Figure. 4.2 : Architecture du 1D CNN

4.4 Expériences

Nous avons fait des expériences avec nos deux modèles en utilisant presque toute la quantité de données du jeu de données. Nous avons formé et testé le modèle CNN contre chaque catégorie d'attaque séparément. Dans le cas de la détection d'anomalie, nous avons formé et validé le DAE avec le trafic normal avant de le tester avec des trafics anormaux. Cette section décrit les différentes étapes que nous avons suivies pour entraîner et tester nos modèles et présente les résultats de nos expérimentations.

4.4.1 EDA

Nous avons conduit une EDA pour explorer le jeu de données en termes de forme et de contenu. La principale motivation de cette analyse est de répondre aux questions suivantes :

- Quelle est la forme des données ?
- Quel type de données avons-nous trouvé sur notre ensemble de données ?
- Comment sont-ils distribués ?
- Quelles sont les caractéristiques ?
- Comment sont-elles corrélées ?

Le CSE-CIC a utilisé 50 machines pour attaquer une organisation victime qui compte 5 départements et comprend 420 machines et 30 serveurs. Ce réseau de victimes est très proche des réseaux d'organisations réelles. L'ensemble de données de +450 Go est constitué du trafic réseau et des captures de journaux système de chaque machine [43]. Parallèlement aux données brutes, Le CICFlowMeter-V3 [43] est utilisé pour extraire 84 caractéristiques du trafic capturé. Nous avons alors obtenu un ensemble de 10 sous-ensembles, capturés en

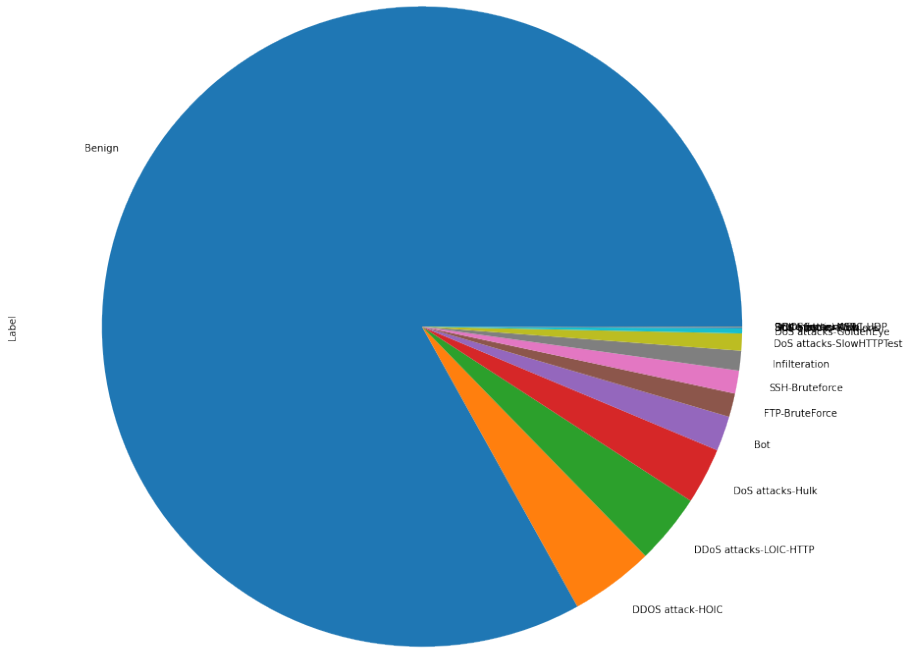


Figure. 4.3 : Distribution du trafic dans CSE-CIC-IDS2018

10 jours différents, composés de 16 232 943 échantillons et 84 caractéristiques (figure 4.3). Parmi ces 84 caractéristiques, seules 80 sont communes à l'ensemble des 10 sous-ensembles, c'est pourquoi il est fréquent de voir des pré-traitements proposés supprimer ces 4 dernières caractéristiques. Dans cet ensemble de données, nous trouvons des caractéristiques de données de trafic réseau bénignes et 14 données de trafic d'attaques différentes les plus récentes. Des informations sur les caractéristiques sont fournies dans le tableau 4.1.

Tableau 4.1 – Information sur les caractéristiques CSE-CIC-IDS2018

CSE-CIC-IDS2018	
Caractéristiques	**Information sur les caractéristiques**
0 – 3, 82	Connexions de réseau
4 - 15	Paquets réseau
16 - 21	Flux réseau
22 – 44	Flux de réseau statistique
45 – 62	Contenu des paquets
63 – 66	Paquets de sous-flux
67 – 78	Caractéristiques du trafic
79	Label
80 – 83	ID de flux, adresses IP source et destination

Tableau 4.2 – Le pourcentage de trafic par type d'attaques

CSE-CIC-IDS2018		
Trafic	**Nombre d'échantillons**	**Taux**
Normal	13 484 708	83.07 %
DDoS attack-HOIC	686 012	7.79 %
DDoS attacks-LOIC-HTTP	576 191	
DDoS attacks-LOIC-UDP	1 730	
DoS attacks-Hulk	461 912	4.03 %
Dos attacks-SlowHTTPTest	139 890	
Dos attacks-GoldenEye	41 508	
Dos attacks-Slowloris	10 990	
Botnet	286 191	1.76 %
FTP-BruteForce	193 360	2.35 %
SSH-BruteForce	187 589	
Infiltration	161 934	0.99 %
Brute Force -Web	611	0.006 %
Brute Force -XSS	230	
SQL Injection	87	
Total	16 232 943	

La description complète de chaque caractéristique est répertoriée dans le [43]. La figure 4.3 montre la distributions des différentes attaques dans le jeu de données. Il s'agit d'un ensemble de données très déséquilibré avec environ 13 484 708 échantillons de trafic normal, soit environ 83 %. Les attaques comme les attaques Web et l'infiltration ne sont que quelques échantillons dans l'ensemble de données. Le déséquilibre des données pourrait entraîner un problème de généralisation et les modèles ne seront pas en mesure de détecter ces minorités et auront tendance à prendre en compte principalement les classes qui constituent la majorité des échantillons. Le nombre de flux par type d'attaque est donné dans le tableau 4.2.

Les noms et les informations des caractéristiques indiquent clairement qu'il peut y avoir une corrélation entre les caractéristiques. Pour le vérifier, nous avons calculé la corrélation entre les caractéristiques et l'avons tracé avec une carte thermique de Seaborn. La figure 4.4 illustre clairement que de nombreuses caractéristiques sont corrélées ; en effet, il existe de fortes corrélations (corrélation supérieure à 0,8). Il est évident que la longueur minimale du paquet dépend du protocole, ou que la longueur de l'en-tête du paquet dépend du protocole et de la longueur totale du paquet, etc. L'analyse de corrélation est effectuée pour vérifier les caractéristiques redondantes et les éliminer. Une analyse plus approfondie est effectuée dans [44] sur l'ensemble de données pour déterminer quelles caractéristiques influencent le plus les processus de formation et d'évaluation. Il faut se référer à l'étude dans [44] pour plus de détails.

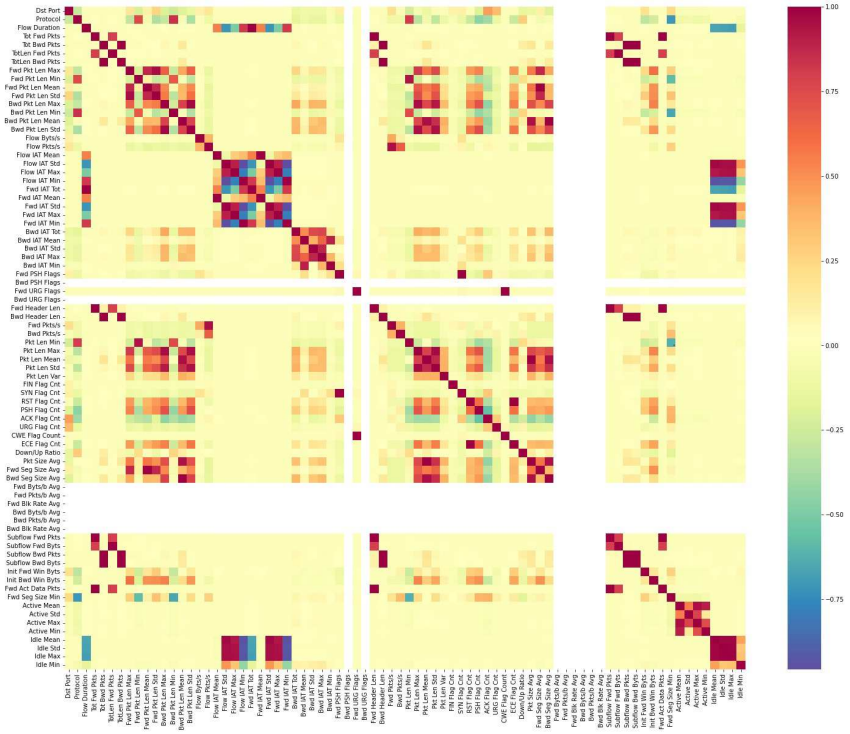


Figure. 4.4 : Corrélation des caractéristiques du CSE-CIC-IDS2018

4.4.2 Pré-traitement des données

Dans notre étape EDA, nous avons vérifié également le pourcentage de NaN et les valeurs dupliquées. L'ensemble de données contient moins de 1 % de NaN et 3,48 % de valeurs dupliquées. Nous avons commencé le pré-traitement en supprimant les NaN et les valeurs dupliquées, car elles représentent une très petite partie, puis nous avons remplacé les valeurs infinies par des NaN avant de les supprimer. Nous avons encodé les étiquettes à l'aide de la fonction Label Encoder et séparé les données par étiquette. À ce stade, nous avons des pré-traitements différents pour chaque modèle.

- Pour le DAE, nous avons isolé le trafic d'attaque à des fins de test et divisé le trafic normal en 50 % d'ensembles d'entraînement/validation et 50 % d'ensemble de tests. Les labels ne sont pas utilisés ici, car nous avons opté pour une détection d'anomalies non supervisée. L'IDS basé sur le DAE est décrit dans la figure 4.5.
- Pour le modèle CNN, nous avons effectué un suréchantillonnage aléatoire pour obtenir une distribution égale pour toutes les classes, car nous savons que les modèles DL sont vulnérables aux données avec des classes déséquilibrées. Nous avons divisé les données en ensembles d'entraînement (60 %), de validation (20 %) et de tests (20 %). Nous avons aussi effectué une extraction des labels pour les convertir en label categorical. L'IDS basé sur le CNN est décrit dans la figure 4.6.

Après ces étapes, nous avons créé un pipeline pour normaliser les données avec la fonction Normalizer et les mettre à l'échelle, entre 0 et 1, avec la fonction MinMaxScaler. L'utilité

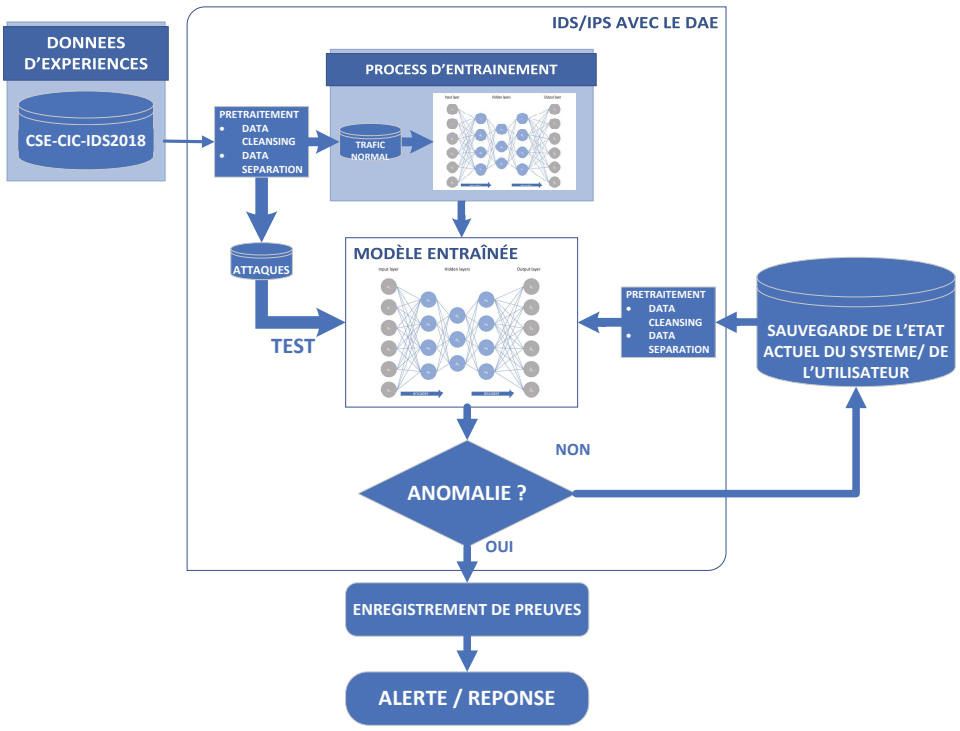


Figure. 4.5 : Processus de l'expérience de la détection d'anomalies

de la normalisation est bien expliquée dans l'article de Jordan [45]. Nous avons finalement remodelé les données au format (données, 1) après la normalisation des données pour le modèle CNN. Nous avons formé notre modèle avec 78 caractéristiques sur 80. La caractéristique de l'horodatage et les labels sont supprimés pour une bonne généralisation.

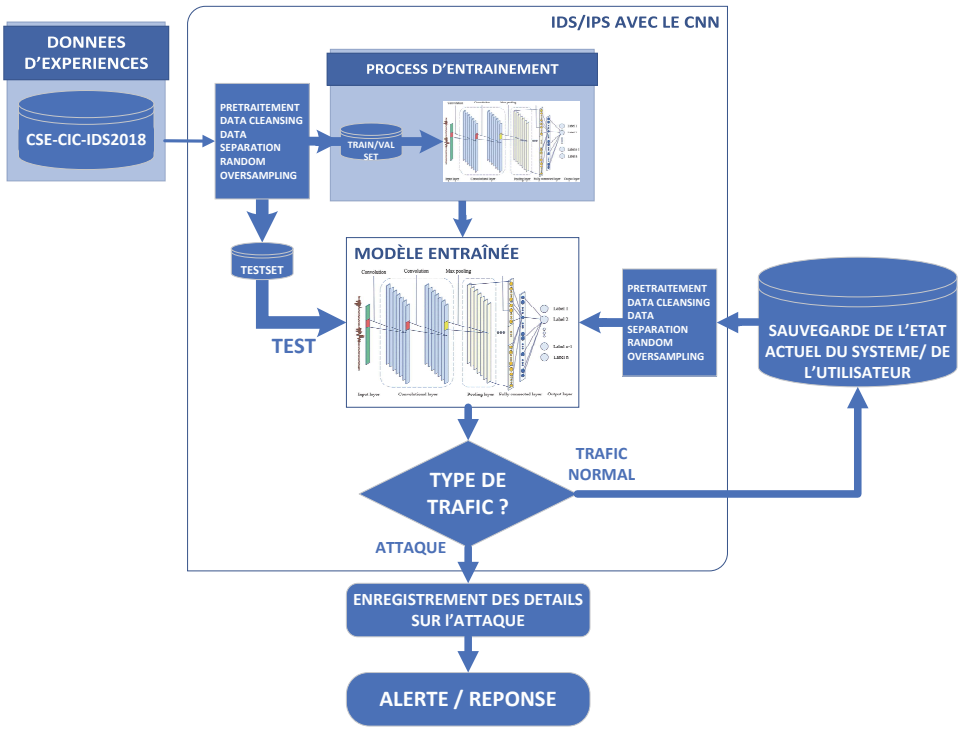


Figure. 4.6 : Processus de l'expérience de la classification d'attaques

4.4.3 Résultats

4.4.3.1 Environnement d'expériences

Nous avons effectué nos expériences en utilisant TensorFlow avec Keras, NumPy, Panda, Scikit Learn et le langage Python3. Notre environnement est un HP EliteBook 840 G1 avec un processeur Core™ i5-4310U (4x 2,6 GHz), Ram 12 Go + 60 Go alloués de manière fixe à partir du disque dur en tant que paging file et un disque dur SSD de 160 Go. Nous avons effectué des expériences de 50 itérations, avec un mécanisme d'Earlystopping avec une patience à 5 pour arrêter l'entraînement en cas d'overfitting et avec des Batch Size de 32, 128, 256 et 512 pour trouver les meilleurs hyperparamètres car nos ressources sont insuffisantes pour les trouver automatiquement avec un Keras Tuner.

4.4.3.2 Métriques d'évaluation

L'objectif de cette étude est de trouver un moyen de minimiser le FAR et d'augmenter le taux de détection en réduisant le FNR. Dans les expériences de classification du trafic, nous avons utilisé des métriques comme la Precision, le Recall, le F1-Score (équations 3.2, 3.3 et 3.4) qui donnent plus de détails sur les faux positifs et les faux négatifs que l'efficacité ou Accuracy.

4.4.3.3 DAE vs attacks : une détection d'anomalies

Le DAE détermine un seuil entre les données normales et les données anormales. La figure 4.7 montre l'évolution de l'erreur de reconstruction au cours des itérations. Notre fonction de prédiction calcule l'erreur de reconstruction et positionne les données reconstruites par rapport au seuil. S'ils sont supérieurs au seuil, ils sont alors détectés comme anomalies. Nous avons calculé un seuil à paramètre variable «n» selon l'équation (4.1). Nous obtenons des résultats de détection d'anomalie DAE pour 4 seuils différents comme indiqué dans le tableau 4.3. Le DAE détecte presque toutes les attaques avec un pourcentage d'anomalies acceptable, voire très bon, sauf pour les attaques DDoS HOIC et DDoS LOIC UDP avec respectivement 0 % et 1,098 % d'anomalies détectées. Les anomalies détectées de DDoS HOIC augmentent à 75,503 % et 100 % avec les autres seuils, mais l'attaque DDoS LOIC UDP reste non détectable pour le DAE. La détection DoS Slowloris présente également un faible pourcentage d'anomalies détectées (maximum 28,128 %). Le DAE détecte également 30,572 % d'anomalies dans le trafic normal utilisé dans le processus de test, comme indiqué dans le tableau 4.3.

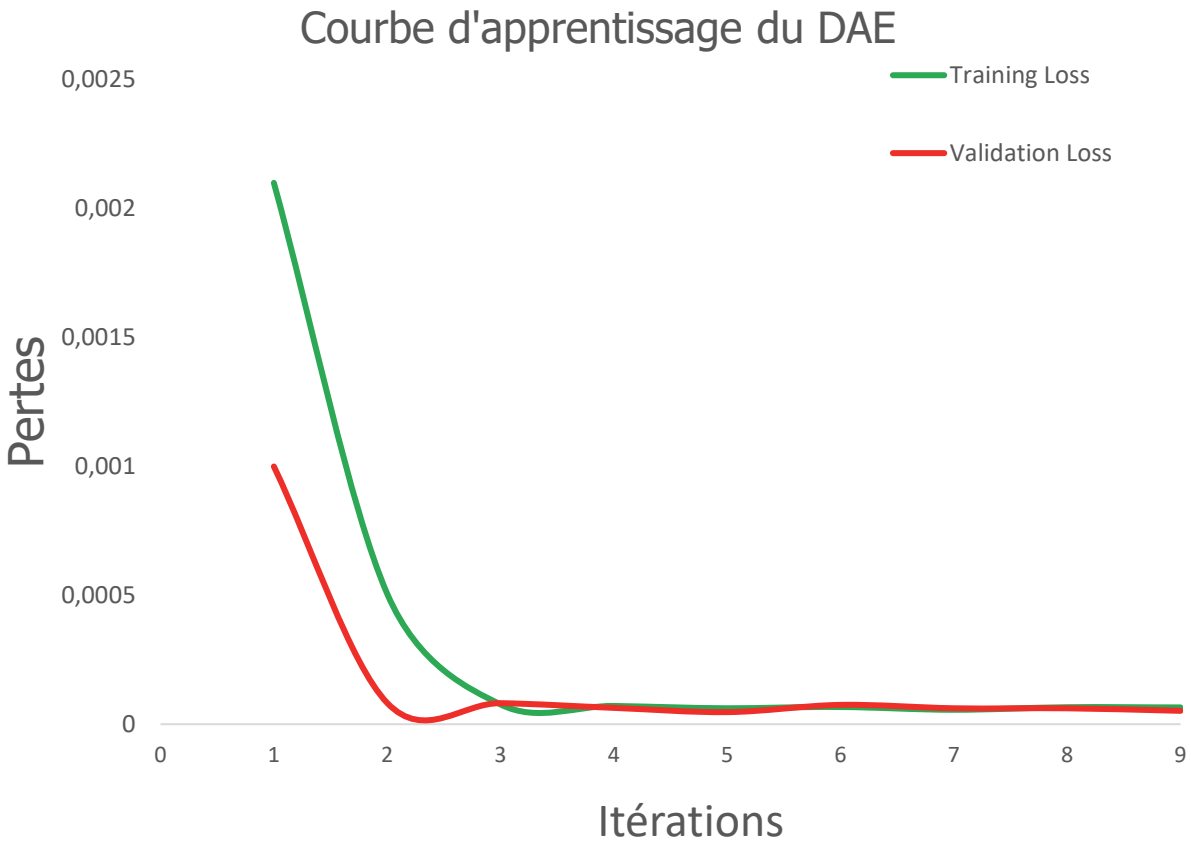


Figure. 4.7 : Courbe d'apprentissage du DAE

Tableau 4.3 – Pourcentages d'anomalies par seuil

Trafic	**Seuil n=1**	**Seuil n=2**	**Seuil n=3**	**Seuil max loss**
Normal test	32.572 %	32.572 %	32.572 %	100 %
Bot	99.998 %	100 %	100 %	100 %
Brute Force Web	99.836 %	99.836 %	100 %	100 %
Brute Force XSS	100 %	100 %	100 %	100 %
DDoS HOIC	0 %	75.503 %	75.503 %	100 %
DDoS LOIC UDP	1.098 %	1.156 %	1.156 %	8.035 %
DDoS LOIC HTTP	99.876 %	99.878 %	99.878 %	99.894 %
DoS GoldenEye	98.280 %	98.664 %	98.794 %	100 %
DoS Hulk	97.434 %	97.447 %	99.474 %	100 %
DoS SlowHTTPTest	68.970 %	83.121 %	97.276 %	100 %
DoS Slowloris	25.318 %	26, 991 %	28.128 %	100 %
FTP brute Force	58.265 %	70.386 %	85.193 %	100 %
Infiltration	97.110 %	97.841 %	98.241 %	100 %
SQL injection	94.252 %	94.253 %	100 %	100 %
SSH brute Force	79.974 %	79.975 %	99.818 %	100 %

4.4.3.4 CNN vs attacks : une classification de trafic

Plutôt que d'entraîner le modèle CNN avec toutes les données, nous avons trouvé plus intéressant d'étudier le comportement du modèle CNN par type d'attaque. Les résultats montrent que le modèle CNN est très efficace avec de très bonnes performances contre les botnets, les attaques DDoS, les attaques DoS, les attaques FTP et SSH BruteForce. Contre ces attaques, les résultats montrent des FAR et FNR très faibles. La figure 4.8 donne les matrices de confusion par type d'attaque avec plus de détails sur la classification. D'autre part, le CNN présente de performances bonnes (92,75 % de Precision, 92,25 % de Recall) contre les attaques Web avec un FAR de 1,225 % (et un nombre élevé d'erreurs de classification des attaques, comme illustré à la figure 4.8), mais un FNR acceptable. Le modèle a particulièrement peiné contre les attaques par infiltration avec un Recall très faible de 28 %. Plus de détails sur les résultats sont donnés dans le tableau 4.4.

Tableau 4.4 – Rapport de classification avec le CNN pour chaque type d'attaque

Trafic	**Rapport de classification**				
	Precision	**Recall**	**F1-score**	**FAR**	**FNR**
Normal	100 %	100 %	100 %	0.018 %	0.122 %
Bot	100 %	100 %	100 %		
Normal	100 %	99 %	99 %	1.225 %	0.081 %
Brute Force Web	86 %	86 %	86 %		
Brute Force XSS	97 %	85 %	90 %		
SQL injection	88 %	99 %	93 %		
Normal	100 %	100 %	100 %	0.336 %	0.005 %
DDoS HOIC	100 %	100 %	100 %		
DDoS LOIC UDP	100 %	100 %	100 %		
DDoS LOIC HTTP	100 %	100 %	100 %		
Normal	100 %	100 %	100 %	0.018 %	0 %
DoS GoldenEye	100 %	100 %	100 %		
DoS Hulk	100 %	100 %	100 %		
DoS SlowHTTPTest	100 %	100 %	100 %		
DoS Slowloris	100 %	100 %	100 %		
Normal	100 %	100 %	100 %	0.022 %	0.003 %
FTP Brute Force	100 %	100 %	100 %		
SSH Brute Force	100 %	100 %	100 %		
Normal	54 %	87 %	67 %	13.070 %	72.433 %
Infiltration	68 %	28 %	39 %		

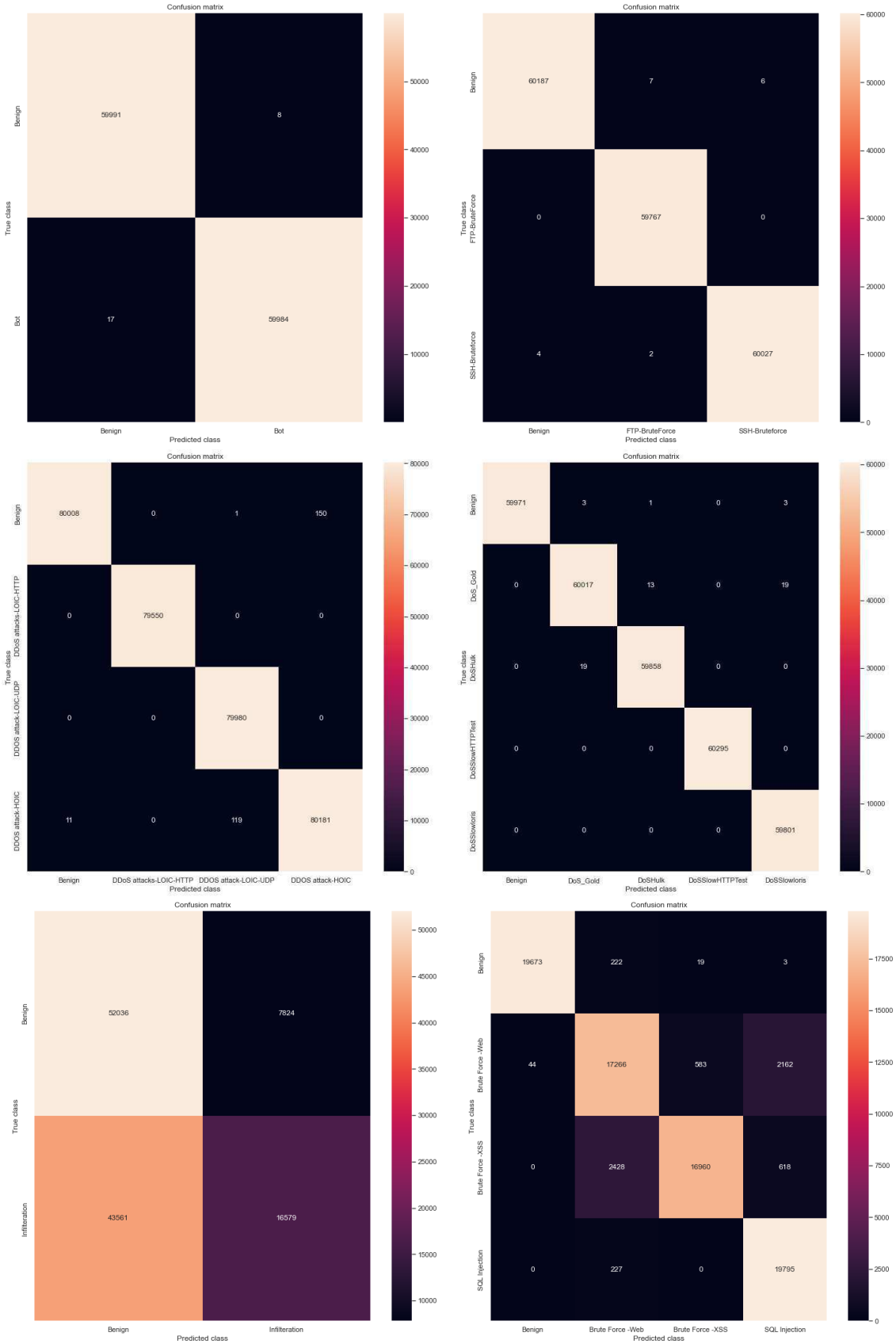


Figure. 4.8 : Matrices de confusion du CNN par type d'attaque

4.5 Discussion

Dans le domaine de la cybersécurité, le Recall et la Precision sont importants, car nous avons besoin d'un Recall aussi élevé que possible, mais nous avons également besoin d'une grande Precision, car nous ne voulons pas d'une charge de travail énorme pour essayer de trouver des attaques fantômes. C'est pourquoi les chercheurs en cybersécurité recherchent un bon F1-score dans les IDS basés sur l'IA.

La détection d'anomalies dans nos expériences est une question de seuil comme nous pouvons le voir dans les résultats du tableau 4.3. Le modèle doit définir le bon seuil pour éviter un FAR et un FNR élevés. Nous avons effectué plusieurs expériences en faisant varier les valeurs n de 1 à 3 et en fixant le seuil au maximum de la perte d'entraînement. Le pourcentage d'anomalies détectées du trafic d'attaque augmente effectivement. Ainsi, les seuils lorsque n = 2 et n = 3, sont plus optimal pour détecter les attaques. Les pourcentages d'anomalies détectées DDoS LOIC UDP et DoS Slowloris restent faibles avec les différents seuils. Avec le dernier seuil, le DAE détecte efficacement toutes les anomalies sauf DDoS LOIC UDP, mais déclenche de nombreuses alarmes, car il détecte tout le trafic normal en test comme des anomalies, ce qui est normal, car les caractéristiques de ce trafic normal n'ont pas forcément la même représentation que le trafic normal dans la phase d'entraînement. Le DoS Slowloris utilise des requêtes plus lentes et partielles pour utiliser les ressources du serveur qui ne pourra libérer aucune des connexions partielles ouvertes. Dans cette attaque, l'attaquant tente d'imiter au maximum un utilisateur normal, c'est pourquoi le DAE a du mal à le détecter. L'attaque DDoS LOIC UDP est une inondation de paquets UDP par des milliers d'utilisateurs coordonnés (par exemple, un botnet) sur la même victime. Étant donné que le trafic d'un seul attaquant dans cette attaque est «normal», cela expliquerait pourquoi le DAE ne peut pas le détecter.

En plus du FAR, l'expérience de classification du trafic décrite plus haut montre quelques erreurs de classification d'attaque, car le modèle a fait quelques erreurs en prédisant une attaque comme une autre. Ce genre d'erreur n'est pas grave, car cela pourrait de toute façon conduire à une bonne alarme, car une attaque est une attaque. Ces expériences nous ont permis de voir que les réseaux CNN sont plus sensibles à certaines attaques qu'à d'autres (voir tableau 4.4). Les expériences pilotées de cette manière sont plus proches de la réalité que de prendre toutes les données pour entraîner le modèle. Dans un scénario d'attaque spécifique, étant donné que du trafic normal est utilisé pour masquer l'attaque, ce trafic est corrélé à cette attaque. Le trafic normal utilisé pour masquer une attaque web peut être différent d'un trafic normal lors d'une attaque DDoS par exemple en termes de caractéristiques. En plus, il serait rare qu'un attaquant exécute plusieurs types d'attaques différents en même temps sur la même victime. C'est pourquoi nous trouvons plus intéressant d'entraîner les modèles par type d'attaque. En effet, nous

trouvons très intéressants à implémenter dans les IDSes, des modèles entraînés par type d'attaque. Les modèles CNN entraînés avec un type d'attaque spécifique sont plus précis. Pour illustrer cela, nous décidons d'entraîner le même modèle CNN avec l'ensemble de données. Les résultats présentés dans le tableau 4.5 montrent que les modèles entraînés avec l'ensemble des données présentent de moins bonnes performances en termes de F1-score et de FAR. L'attaque par infiltration brouille le modèle qui fait plus d'erreurs dans la classification du trafic normal que les modèles précédents. Cela affecte également la classification des attaques DoS, comme illustré à la figure 4.9.

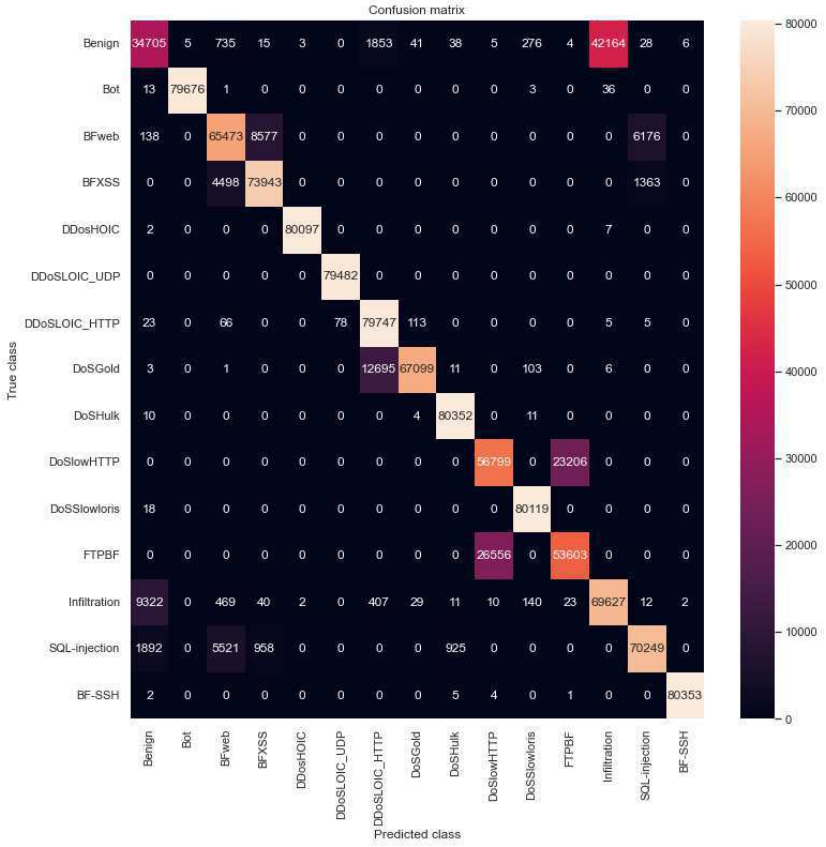


Figure. 4.9 : Matrice de confusion du CNN vs toutes les attaques combinées

Les deux modèles peuvent gérer des données inconnues, car ils présentent des performances prometteuses en matière de validation et de prédiction. Les attaques zero-day pourraient être détectées d'une manière acceptable. Comparé au DAE, le modèle CNN donne plus de détails sur la classification et peut faciliter la vie des administrateurs de réseaux, car le DAE ne déclenche des alarmes que lorsque des anomalies sont détectées. En comparaison, notre CNN détecte facilement certaines attaques comme DDoS LOIC UDP et Slowloris là où le DAE a vraiment du mal. D'autre part, notre DAE détecte efficacement les attaques Web et les infiltrations là où notre CNN peine. Chaque modèle DL a sa propre façon de représenter les caractéristiques du trafic ce qui explique la diversité dans les capacités de détection. Une combinaison de ces deux modèles pourrait être intéressante.

Tableau 4.5 – Rapport de classification avec le CNN contre CSE-CIC-IDS2018

Trafic	**Rapport de classification**				
	Precision	**Recall**	**F1-score**	**FAR**	**FNR**
Normal	75 %	43 %	55 %	79.818 %	1.111 %
Bot	100 %	100 %	100 %		
Brute Force Web	85 %	81 %	83 %		
Brute Force XSS	89 %	93 %	91 %		
DDoS HOIC	100 %	100 %	100 %		
DDoS LOIC UDP	100 %	100 %	100 %		
DDoS LOIC HTTP	84 %	100 %	91 %		
DoS GoldenEye	100 %	84 %	91 %		
DoS Hulk	99 %	100 %	99 %		
DoS SlowHTTPTest	68 %	71 %	70 %		
DoS Slowloris	99 %	100 %	100 %		
FTP Brute Force	70 %	67 %	68 %		
Infiltration	62 %	87 %	73 %		
SQL injection	90 %	88 %	89 %		
SSH Brute Force	100 %	100 %	100 %		

4.6 Conclusion

Nous avons proposé dans ce chapitre deux modèles dont un modèle de classification par un CNN et un autre de détection d'anomalies par un DAE. Nous avons exploré la partie traitée de l'ensemble de données CSE-CIC-IDS2018. Cela nous a aidé dans l'étude à bien exploiter les données, en pré-traitant l'ensemble de données pour former nos modèles. Les deux modèles ont bien les données du CSE-CIC-IDS2018, les résultats ont montré que le DAE et le modèle CNN peuvent détecter efficacement presque toutes les attaques. En présentant des très bonnes performances sur l'ensemble de validation et l'ensemble de tests, ils ont montré des capacités à détecter les attaques zero-day à quelques exceptions près. En effet, certaines attaques comme l'infiltration et DDoS LOIC UDP restent furtives respectivement pour le CNN et le DAE, car elles ont des propriétés de trafic bénignes. Le CNN formé par type d'attaque a présenté de meilleures performances que le CNN formé avec toutes les données. Les résultats du CNN par type d'attaque ont montré une amélioration en termes de FAR et de FNR. Nous avons voulu illustrer dans cette étude que les modèles de Deep learning ont la capacité de détecter les cyberattaques connues et inconnues mais aussi qu'un seul modèle n'est pas efficace pour détecter toutes les attaques, et ne peut être implémenté seul sur un IDS. Au lieu d'utiliser un grand modèle avec des millions de paramètres, nous pouvons utiliser plusieurs petits modèles chacun dans un but spécifique. Le véritable défi de notre modèle DAE est le FAR élevé à réduire davantage.

Conclusion générale

Dans ce monde maintenant connecté, nous développons beaucoup de nos activités en ligne. Les individus de même que les organisations se basent sur internet et la technologie pour faciliter leurs opérations. Ceux-ci ont apporté un lot d'avantages ainsi qu'un contre coup. En effet, en plus de notre négligence, nous sommes actuellement exposées à des risques provenant d'internet. Les cybercriminels tentent constamment de trouver une ouverture vers nos systèmes pour compromettre la sécurité de ceux-ci.

L'évolution technologique nous permet de nous défendre, mais permet aussi aux cybercriminels d'améliorer leurs techniques jusqu'à parvenir à rendre nos défenses inefficaces avec les attaques zero-day. Nous avons étudié la panoplie de techniques et technologies utilisées pour mettre en oeuvre les contre-mesures. Étant donné que les impacts des attaques zero-day ont un caractère inconnu, se prémunir d'une solution de prévention ou d'atténuation est d'actualité.

Nous avons étudié les IDS qui ont donné une petite lueur d'espoir quant à la détection de l'inconnu à faible coût et avec une expertise « uniquement » requise à leur conception contrairement aux SIEM et aux services de Threat intelligence.

Dans le but d'améliorer la détection des IDS, en remplaçant le module d'analyse, nous avons proposé deux modèles de Deep Learning aux processus d'apprentissage et de prédiction très différents, mais aux résultats similaires et prometteurs. En effet, nos résultats ont montré que notre modèle de CNN utilisé pour une classification des attaques et notre modèle DAE utilisé pour une détection d'anomalies peuvent détecter les attaques inconnues. En plus, ils détectent les attaques utilisées dans nos expériences avec un faible taux d'erreurs (FPR et FNR). Les modèles ont été entraînés, évalués et testés avec les trafics de l'ensemble de données CSE-CIC-IDS2018. En bonus, notre modèle de classification fournit des détails sur les attaques. Les points à améliorer dans notre étude sont :

- La prise en charge de classes de trafic inconnues (unknown unknown) ;
- Un meilleur pré-traitement de données ;
- La réduction du temps de détection qui dépend du temps de prédiction.

L'étude d'autres modèles et les combinaisons de modèles sont aussi des perspectives à explorer.

Bibliographie

[1] Liu Hua Yeo, Xiangdong Che, and Shalini Lakkaraju. Modern intrusion detection systems. *CoRR*, abs/1708.07174, 2017. URL http://arxiv.org/abs/1708.07174.

[2] Khraisat Ansam, Gondal Iqbal, Vamplew Peter, and Kamruzzaman Joarder. Survey of intrusion detection systems : techniques, datasets and challenges. *Cybersecurity*, 2(1) :20, jul 2019. doi : 10.1186/s42400-019-0038-7. URL https://doi.org/10.1186/s42400-019-0038-7.

[3] Jan Lansky, Saqib Ali, Mokhtar Mohammadi, Mohammed Kamal Majeed, Sarkhel H. Taher Karim, Shima Rashidi, Mehdi Hosseinzadeh, and Amir Masoud Rahmani. Deep learning-based intrusion detection systems : A systematic review. *IEEE Access*, 9 :101574–101599, 2021. doi : 10.1109/ACCESS.2021.3097247.

[4] Arwa Aldweesh, Abdelouahid Derhab, and Ahmed Z. Emam. Deep learning approaches for anomaly-based intrusion detection systems : A survey, taxonomy, and open issues. *Knowl. Based Syst.*, 189, 2020. doi : 10.1016/j.knosys.2019.105124. URL https://doi.org/10.1016/j.knosys.2019.105124.

[5] ISO and IEC. Information technology—security techniques-information security risk management. *ISO/IEC FIDIS 27005:2008*, 2008.

[6] Internet Engineering Task Force IETF. Rfc 4949 internet security glossary, version 2. *RFC 4949*, 2007.

[7] Dani Deahl. Verizon partner data breach exposes millions of customer records. https://www.theverge.com/2017/7/12/15962520/verizon-nice-systems-data-breach-exposes-millions-customer-records, jul 2017.

[8] Frans Rosén. How patreon got hacked - publicly exposed werkzeug debugger. https://labs.detectify.com/2015/10/02/how-patreon-got-hacked-publicly-exposed-werkzeug-debugger/, oct 2015.

[9] United States Government Accountability Office US GAO. Actions taken by equifax and federal agencies in response to the 2017 breach, 2018.

[10] CISCO TALOS. Cisco talos - vulnerability information. https://talosintelligence.com/vulnerability_info, jun 2022.

[11] Jim Finkle. Bangladesh bank hackers compromised swift software, warning issued. https://www.reuters.com/article/us-usa-nyfed-bangladesh-malware-exclusiv-idUSKCN0XM0DR, apr 2016.

[12] Federal Trade Commission FTC. Equifax to pay $575 million as part of settlement with ftc, cfpb, and states related to 2017 data breach. https://www.ftc.gov/news-events/press-releases/2019/07/equifax-pay-575-million-part-settlement-ftc-cfpb-states-related, jul 2019.

[13] Iyengar Rishi and CNN-Business. Uber's former security chief charged for allegedly concealing data breach. https://edition.cnn.com/2020/08/20/tech/uber-security-chief-charged/index.html, aug 2020.

[14] Marius POPA. Analysis of zero-day vulnerabilities in java. *Journal of Mobile, Embedded and Distributed Systems*, 5(3), 2013.

[15] Aditya K Sood and Richard Enbody. *Chapter 4 - System Exploitation*, pages 37–75. Syngress, 2014.

[16] Fortinet. Types of cyber attacks. https://www.fortinet.com/resources/cyberglossary/types-of-cyber-attacks, 2022.

[17] Nektaria Kaloudi and Jingyue Li. The ai-based cyber threat landscape : A survey. *ACM Comput. Surv.*, 53(1) :20:1–20:34, 2020. doi : 10.1145/3372823. URL https://doi.org/10.1145/3372823.

[18] Mr. Dustin Fraze. Cyber grand challenge (cgc). https://www.darpa.mil/program/cyber-grand-challenge, Aug 2016.

[19] MANDIANT. Mandiant apt. https://www.mandiant.com/search?search=APT, 2022.

[20] Corporation THE-MITREE. Enterprise tactics. https://attack.mitre.org/techniques/T1190/, 2022.

[21] Dan McWhorter. Apt1 exposing one of china's cyber espionage units, 2013.

[22] Edward Humphreys. *Implementing the ISO/IEC 27001 ISMS Standard*. ARTECH HOUSE, 2016. ISBN 13:978-1-60807-930-8.

[23] Darril Gibson. *Comptia Security + : Get Certified Get Ahead SY0-501 Study Guide*. YCDA-LLC, 2017.

[24] Professor Messer. Risk analysis - comptia security+ 5.4. https://www.professormesser.com/security-plus/sy0-601/sy0-601-video/risk-analysis/, 2021.

[25] Philippe Biondi. *Vers un environnement d'objets connectés efficient et sécurisé grâce à l'analyse de trafic et l'apprentissage automatique*. PhD thesis, Unknown, 2001.

[26] Huawei Technologies Co. Huawei. *HCIP Network Security Device and Networking*. Huawei Technologies Co., 2017.

[27] The Snort-Project. Snort 2.9.16 users manual. https://www.snort.org/, 2020.

[28] Avewe Bassene. *Vers un environnement d'objets connectés efficient et sécurisé grâce à l'analyse de trafic et l'apprentissage automatique*. PhD thesis, Universite Cheikh Anta Diop de Dakar, 2022.

[29] NIST. National vulnerability database. https://nvd.nist.gov/, 2022.

[30] David Chismon and Martin Ruks. Threat intelligence : Collecting, analysing, evaluating, 2015.

[31] Yaman Roumani. Patching zero-day vulnerabilities : an empirical analysis. *J. Cybersecur.*, 7(1), 2021. doi : 10.1093/cybsec/tyab023. URL https://doi.org/10.1093/cybsec/tyab023.

[32] Umesh Kumar Singh, Chanchala Joshi, and Dimitris Kanellopoulos. A framework for zero-day vulnerabilities detection and prioritization. *J. Inf. Secur. Appl.*, 46 : 164–172, 2019. doi : 10.1016/j.jisa.2019.03.011. URL https://doi.org/10.1016/j.jisa.2019.03.011.

[33] Jiyeon Kim, Yulim Shin, and Eunjung Choi. An intrusion detection model based on a convolutional neural network. *Journal of Multimedia Information System*, 6 (4) :165–172, 2019. doi : 10.33851/JMIS.2019.6.4.165. URL https://doi.org/10.33851/JMIS.2019.6.4.165.

[34] Li Yong and Zhang Bo. An intrusion detection model based on multi-scale cnn. In *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pages 214–218, 2019. doi : 10.1109/ITNEC.2019.8729261.

[35] Riaz Ullah Khan, Xiaosong Zhang, Mamoun Alazab, and Rajesh Kumar. An improved convolutional neural network model for intrusion detection in networks. In *2019 Cybersecurity and Cyberforensics Conference (CCC)*, pages 74–77, 2019. doi : 10.1109/CCC.2019.000-6.

[36] Meliboev Azizjon, Alikhanov Jumabek, and Wooseong Kim. 1d cnn based network intrusion detection with normalization on imbalanced data. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 218–224, 2020. doi : 10.1109/ICAIIC48513.2020.9064976.

[37] Jiyeon Kim, Jiwon Kim, Hyunjung Kim, Minsun Shim, and Eunjung Choi. Cnn-based network intrusion detection against denial-of-service attacks. *Electronics*, 9 (6), 2020. ISSN 2079-9292. doi : 10.3390/electronics9060916. URL https://www.mdpi.com/2079-9292/9/6/916.

[38] Jihoon Yoo, Byeongjun Min, Sangsoo Kim, Dongil Shin, and Dongkyoo Shin. Study on network intrusion detection method using discrete pre-processing method and convolution neural network. *IEEE Access*, 9 :142348–142361, 2021. doi : 10.1109/ACCESS.2021.3120839.

[39] Yunxiang Kang, Minsheng Tan, Ding Lin, and Zhiguo Zhao. Intrusion detection model based on autoencoder and XGBoost. *Journal of Physics : Conference Series*, 2171(1) :012053, jan 2022. doi : 10.1088/1742-6596/2171/1/012053. URL https://doi.org/10.1088/1742-6596/2171/1/012053.

[40] Marta Catillo, Massimiliano Rak, and Umberto Villano. Discovery of dos attacks by the ZED-IDS anomaly detector. *J. High Speed Networks*, 25(4) :349–365, 2019. doi : 10.3233/JHS-190620. URL https://doi.org/10.3233/JHS-190620.

[41] Rak Massimiliano Catillo Marta and Villano Umberto. 2l-zed-ids : A two-level anomaly detector for multiple attack classes. In Leonard Barolli, Flora Amato, Francesco Moscato, Tomoya Enokido, and Makoto Takizawa, editors, *Web, Artificial Intelligence and Network Applications*, pages 687–696, Cham, 2020. Springer International Publishing. ISBN 978-3-030-44038-1.

[42] Jeremy Jordan. Introduction to autoencoders. https://www.jeremyjordan.me/autoencoders/, mar 2018.

[43] CSE-CIC. Cse-cic-ids2018 on aws. https://www.unb.ca/cic/datasets/ids-2018.html, 2018.

[44] CSTUB-EDA. https://github.com/cstub/ml-ids/tree/master/notebooks/02_exploratory-data-analysis, 2019.

[45] Jeremy Jordan. Normalizing your data (specifically, input and batch normalization). https://www.jeremyjordan.me/batch-normalization/, 2018.