الجـــــمـــهـــوريـــــــة الـجـزائـريـــة الديمـــقـراطـيـــة الـشـــعـبـيـــة

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التـــــعـــليـــم العـــــــــالي والبـــــــحث العـــــــــلـــمـــي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جـــــــــــــــــــــــامعة سعيدة – د. الطاهر مولاي –

Université Saïda – Dr. Tahar Moulay –

Faculté des Mathématiques, de l'Informatique et des Télécommunications

MEMOIRE

Présenté pour l'obtention du **Diplôme** de **MASTER en** Télécommunications

Spécialité : Réseaux et Télécommunications

Par : M. ATTOU Lahcene

M. BOUHALOUANE Abdelatif

Détection des cyber-attaques dans les systèmes à infrastructures critiques par des modèles de Deep Learning hybrides

Soutenu, le 13 /06/ 2026, devant le jury composé de :

M. MOKADEM Djelloul	MCB	Président
M. BOUYEDDOU Benamar	MCA	Rapporteur
M. TAMI Abdelkader	MCA	Examinateur

2025 / 2026

Remerciements

À notre directeur de recherche, Monsieur Bouyeddou Benamar :

Nous tenons à adresser nos sincères remerciements à notre encadrant pour son soutien indéfectible, sa rigueur méthodologique et la confiance qu'il nous a accordée. Ses orientations éclairées, sa maîtrise scientifique et sa vision critique ont constitué le socle de notre démarche. Grâce à son précieux encadrement, nous avons pu structurer notre pensée et valoriser pleinement ce travail de recherche.

À Monsieur Mokadem , membre du jury :

Nous adressons nos chaleureux remerciements aux membres du jury pour l'intérêt qu'ils ont porté à notre travail à travers une lecture attentive et rigoureuse. Leurs remarques pertinentes, leurs critiques constructives et leurs questions éclairées ont grandement contribué à enrichir la valeur de ce mémoire, et leur expertise scientifique nous a été particulièrement précieuse.

À Monsieur Tami, membre du jury :

C'est un honneur pour nous d'exprimer notre profonde gratitude aux membres du jury pour avoir examiné notre travail avec autant de rigueur et de bienveillance. Leurs précieuses orientations, leurs critiques constructives et les discussions enrichissantes lors de la soutenance ont permis de perfectionner ce mémoire et d'élever sa qualité scientifique.

Ce mémoire est le fruit de l'accompagnement et du soutien de chacune de ces personnes ; qu'elles trouvent ici l'expression de notre plus profonde considération.

En dernier lieu, c'est avec une immense émotion que nous adressons un grand merci à nos familles respectives, **ATTOU** et **BOUHALOUAN**. Vos sacrifices matériels, votre présence réconfortante et votre soutien inconditionnel durant toute cette période de préparation ont constitué le pilier sur lequel nous nous sommes appuyés pour mener à bien ce travail.

Dédicaces

À ceux qui ont éclairé mon chemin et soutenu mes pas...

À mon cher père, Exemple de rigueur et de sacrifice. Ton soutien infaillible et ta fierté silencieuse ont toujours été mon moteur pour aller de l'avant. Que ce travail soit le reflet de ta générosité.

À ma tendre mère, Source d'amour inconditionnel et de patience infinie. Tes prières quotidiennes et ta bienveillance ont été ma plus grande force pour surmonter les doutes. Ce mémoire est avant tout le fruit de tes sacrifices.

À mes frères et sœurs, Mes piliers, mes compagnons de route et mon refuge. Merci pour votre complicité, votre présence rassurante au quotidien et votre soutien moral inébranlable.

À ma petite sœur, Le rayon de soleil de la maison. Ta joie de vivre et tes encouragements m'ont toujours redonné le sourire durant les moments de fatigue.

À toute ma précieuse famille, *Je dédie ce modeste travail en signe de reconnaissance, d'amour et de profonde gratitude.*

Dédicaces

Je dédie ce travail de recherche :

À mes chers parents, mon père et ma mère, pour leur amour inconditionnel, leurs sacrifices infinis et leur soutien indéfectible tout au long de mon parcours universitaire. Aucun mot ne saurait exprimer ma profonde gratitude pour l'éducation et les valeurs qu'ils m'ont transmises.

À ma tendre grand-mère, que Dieu lui prête longue vie et bonne santé. Ses prières quotidiennes et sa bienveillance ont été ma source de force et de bénédiction.

À mes frères, Oussama et Chakib, pour leur complicité, leurs encouragements et leur présence rassurante à mes côtés. Que ce travail soit le reflet de notre fraternité.

À ma chère sœur, Inas, pour sa douceur, son affection et son soutien moral qui m'ont toujours grandement motivé.

À toute ma famille et à ceux qui me sont chers.

Résumé

Les systèmes de classification des cyber-attaques sont largement répandus de nos jours pour la sécurité des systèmes à infrastructures critiques (IC). Diverses techniques de machine learning ont été utilisées pour mettre en place ces systèmes, qui ont été bien propulsés avec l'apparition des technologies de deep learning et leurs succès dans plusieurs domaines.

Dans ce mémoire, nous avons élaboré une approche de classification des cyber-attaques dans les réseaux industriels basée sur cinq modèles hybrides d'apprentissage profond : LSTM-GRU, RNN-GRU, CNN-GRU, RNN-LSTM et CNN-LSTM. La méthodologie mise en place a été validée sous la base de données réelle ICSNAD, contenant quatre types d'attaques : PORSCAN, TEARDROP, ICMPREDIRECT et ARP POISONING.

Les résultats obtenus montrent que l'approche utilisée a permis d'atteindre des performances de classification prometteuses, avec une supériorité du modèle LSTM-GRU atteignant un rappel de 97,34% pour TEARDROP et une précision de 98% pour TEARDROP. L'étude révèle également l'impact crucial du déséquilibre des classes sur les performances de détection.

Mots clés : Infrastructures critiques, IC, Cyber-attaques, Machine learning, Deep Learning, Modèles hybrides, LSTM-GRU, RNN-GRU, CNN-GRU, ICSNAD, base de données, ICSNAD

.

Abstract

Cyber-attack classification systems are widely used nowadays for the security of critical infrastructures systems (IC). Various machine learning techniques have been employed to implement these systems, which have been significantly advanced with the emergence of deep learning technologies and their successes in several fields.

In this dissertation, we developed a cyber-attack classification approach for industrial networks based on five hybrid deep learning models: LSTM-GRU, RNN-GRU, CNN-GRU, RNN-LSTM, and CNN-LSTM. The established methodology was validated using the real-world ICSNAD dataset, containing four types of attacks: PORSCAN, TEARDROP, ICMPREDIRECT, and ARP POISONING.

The obtained results show that the proposed approach achieved promising classification performance, with the LSTM-GRU model outperforming others, reaching a recall of 97.34% for TEARDROP and a precision of 98% for TEARDROP. The study also reveals the crucial impact of class imbalance on detection performance.

Keywords: Critical infrastructure, cyber-attacks, Machine learning, Deep Learning, Hybrid models, LSTM-GRU, RNN-GRU, CNN-GRU, ICSNAD dataset.

الملخص

تنتشر أنظمة تصنيف الهجمات السيبرانية بشكل واسع في الوقت الحاضر لضمان أمن الأنظمة الصناعية .(IC) وقد أُستُخدِمت تقنيات مختلفة من التعلم الآلي لبناء هذه الأنظمة، والتي شهدت دفعة قوية مع ظهور تكنولوجيات التعلم العميق ونجاحاتها في مجالات متعددة.

في هذه المذكرة، قمنا بتطوير نهج لتصنيف الهجمات السيبرانية في الشبكات الصناعية يعتمد على خمسة نماذج هجينة للتعلم العميق، وهي**LSTM-GRU** :، و **RNN-GRU**، و **CNN-GRU**، و **RNN-LSTM**، و .**CNN-LSTM**وقد تم التحقق من صحة المنهجية المتبعة باستخدام قاعدة البيانات الحقيقية **ICSNAD**، والتي تحتوي على أربعة أنواع من الهجمات**PORSCAN** :، و **TEARDROP**، و **ICMPREDIRECT**، و .**ARP POISONING**

تُظهر النتائج التي تم الحصول عليها أن النهج المستخدم قد سمح بتحقيق أداء تصنيف واعد، مع تفوق نموذج **-LSTM GRU**الذي حقق معدل استدعاء (Rappel) بنسبة **97.34%**لهجومTEARDROP ، ودقة (Précision) بنسبة **98%** لنفس الهجوم. كما تكشف الدراسة أيضًا عن التأثير الحاسم لعدم توازن الفئات (Déséquilibre des classes) على أداء الكشف.

الكلمات المفتاحية :البنية التحتية الحيوية الهجمات السيبرانية ، التعلم الآلي، التعلم العميق، النماذج الهجينة، LSTM-GRU، RNN-GRU، CNN-GRU، ICSNAD.

Table des matières

Liste des figures

CHAPITRE I

CHAPITRE II

CHAPITRE III

Liste des Tableaux

CHAPITRE I

CHAPITRE II

CHAPITRE III

Liste des abréviations

CNN : Convolutional Neural Networks

CPL : Courants Porteurs en Ligne

CNC : Couches de Neurones Convolutionnelles

CSIS : Center for Strategic and International Studies

DDoS : Distributed Denial of Service

DHCPv6 : Dynamic Host Configuration Protocol for IPv6

DOS : Denial of Service

FDIA : False Data Injection Attack

GEO : Geostationary Earth Orbit

GRU : Gated Recurrent Unit

ICSNAD : Industrial Control Systems Network Attack Dataset

IDS : Intrusion Detection Systems

IED : Intelligent Electronic Device

IPv6 : Internet Protocol version 6

LPWAN : Low-Power Wide-Area Network

LSTM : Long Short-Term Memory

MIMT : Man-in-the-Middle

MQTT : Message Queuing Telemetry Transport

NIPP : National Infrastructure Protection Plan

PCAP : Packet Capture

RNN : Recurrent Neural Networks

RTU : Remote Terminal Unit

SCADA : Supervisory Control and Data Acquisition

TCP/IP : Transmission Control Protocol / Internet Protocol

TIC : Technologies de l'Information et de la Communication

VSAT : Very Small Aperture Terminal

Introduction Générale

Introduction Générale

De nos jours, les systèmes à infrastructure Critiques (IC) occupent une place prédominante dans tous les secteurs stratégiques : l'énergie, le transport, l'eau, la fabrication, la santé et le domaine militaire. Initialement isolés et conçus pour fonctionner en environnements clos, ces systèmes sont à présent de plus en plus interconnectés aux réseaux informatiques d'entreprise et à l'internet. Ce phénomène, catalysé par l'essor de l'industrie 4.0 et l'Internet des Objets (IoT), offre de nombreux avantages en termes de productivité, de maintenance à distance et d'optimisation des processus. Cependant, cette connectivité accrue expose les systèmes industriels à des cyber-attaques de plus en plus sophistiquées.

Les cyber-attaques contre les systèmes IC sont en forte hausse ces dernières années et représentent aujourd'hui un risque réel qui menace les réseaux industriels, les automates programmables, les systèmes de contrôle-commande (SCADA) et les infrastructures critiques. Contrairement aux systèmes informatiques classiques, une attaque réussie contre un système IC peut avoir des conséquences catastrophiques : dommages matériels, atteintes à l'environnement, interruptions de services essentiels (électricité, eau potable…), voire pertes humaines. Les cyber-attaquants peuvent exploiter les vulnérabilités des protocoles industriels (Modbus, Profibus…) et des équipements pour espionner, modifier ou détruire des processus critiques.

Face à toutes ces menaces, la sécurité optimale de ces systèmes est devenue un enjeu stratégique national. Pour atteindre cet objectif, différents mécanismes de protection ont été développés, tels que les systèmes de détection d'intrusions (IDS), les pare-feu industriels et les solutions de segmentation réseau. Malheureusement, ces outils traditionnels sont dans la plupart du temps inefficaces face aux cyber-attaques avancées, zero-day et aux menaces persistantes, dont la propagation peut s'avérer extrêmement rapide et silencieuse.

En effet, l'utilisation des méthodes classiques de détection basées sur des signatures n'est plus suffisante. De nouvelles technologies de protection avancées demeurent indispensables. De ce fait, les acteurs de la cyber-sécurité IC s'orientent vers l'utilisation de l'apprentissage automatique, un sous-domaine de l'intelligence artificielle (IA) qui a été utilisé avec succès dans plusieurs domaines, comme la reconnaissance d'images, le traitement du langage naturel et la

prise de décision, afin de renforcer l'efficacité de leurs produits face aux problèmes de détection des cyber-attaques.

Dans ce sens, le Deep Learning ou apprentissage profond, sous-domaine du machine learning, se présente comme une série d'approches très prometteuses pour la cyber-sécurité, particulièrement avec la disponibilité croissante de grandes quantités de données issues des réseaux IC (journaux d'événements, captures réseau, historiques de processus). Les méthodologies traditionnelles d'apprentissage machine reposent sur l'ingénierie et la sélection manuelle des caractéristiques, ce qui nécessite une bonne expertise du domaine étudié et des attaques potentielles.

Dans cet axe, l'objectif de ce mémoire est d'investiguer l'utilité des modèles hybrides d'apprentissage profond pour la classification des cyber-attaques dans les réseaux IC. Précisément, nous présentons une méthodologie de classification basée sur cinq architectures hybrides : LSTM-GRU, RNN-GRU, CNN-GRU, RNN-LSTM et CNN-LSTM. Ces modèles combinent les avantages des réseaux convolutionnels (CNN) pour l'extraction des caractéristiques spatiales, des réseaux récurrents (RNN) pour la modélisation des séquences temporelles, et des unités à mémoire (LSTM et GRU) pour la capture des dépendances à long terme. Une analyse approfondie des performances en termes d'Accuracy, Précision, Rappel, et F1-score est menée en utilisant la base de données réelle ICSNAD (Industrial Control Systems Network Attack Dataset) . Nous avons utilisé quatre types d'attaques distincts : PORSCAN, TEARDROP, ICMPREDIRECT et ARP POISONING.

La suite manuscrit s'organise comme suit :

Le premier chapitre présente les principes de base des réseaux IC (architectures, protocoles, vulnérabilités) et les notions générales liées à la problématique des cyber-attaques dans les systèmes IC.

Le deuxième chapitre comprend un état de l'art sur les approches de Deep Learning, avec concentration sur les architectures couramment utilisées (CNN, RNN, LSTM, GRU).

Le troisième chapitre reporte nos résultats de classification pour les quatre types d'attaques, avec une analyse comparative des cinq modèles hybrides.

Chapitre I :

Les systèmes à infrastructures critiques

Chapitre I : Les systèmes à infrastructures critiques

I.1. Introduction

De nos jours, le développement fulgurant de la technologie a conduit à sa pénétration dans tous les secteurs de l'industrie et des infrastructures. Ce phénomène est inévitable, car les sociétés modernes ont des besoins accrus en ressources pour fonctionner correctement. Grâce à l'utilisation intensive de la technologie, la production peut être augmentée et les demandes de la société peuvent être satisfaites. Ces infrastructures sont qualifiées de critiques en raison des rôles essentiels qu'elles jouent dans la continuité des services vitaux tels que l'approvisionnement en électricité, l'eau potable, les transports, les télécommunications ou encore la santé.

Cependant, l'intégration croissante des technologies dans le fonctionnement des infrastructures critiques introduit de nouvelles vulnérabilités. Si cette transformation numérique améliore l'efficacité et la réactivité des systèmes, elle élargit également la surface d'attaque potentielle. En conséquence, des acteurs malveillants peuvent exploiter ces failles pour perturber, endommager ou prendre le contrôle de ces infrastructures, causant ainsi des dommages économiques, environnementaux ou humains considérables. Ce chapitre présente les concepts fondamentaux des systèmes à infrastructures critiques et expose les principaux enjeux de sécurité qui les concernent.

I.2. Définition des Infrastructures Critiques (IC)

Le terme infrastructure critique désigne les actifs physiques et virtuels, les systèmes et les réseaux qui sont essentiels au fonctionnement d'une société. Ces infrastructures sont considérées comme vitales en raison de leur impact significatif sur la sécurité nationale, la stabilité économique, ainsi que sur la santé et la sécurité publiques. Leur altération, destruction ou perturbation peut avoir des conséquences graves, entraînant des vulnérabilités sociétales et des effets néfastes sur les citoyens. L'infrastructure critique est une composante fondamentale de la société moderne, englobant les actifs, les systèmes et les réseaux indispensables au maintien des fonctions sociétales vitales [1].

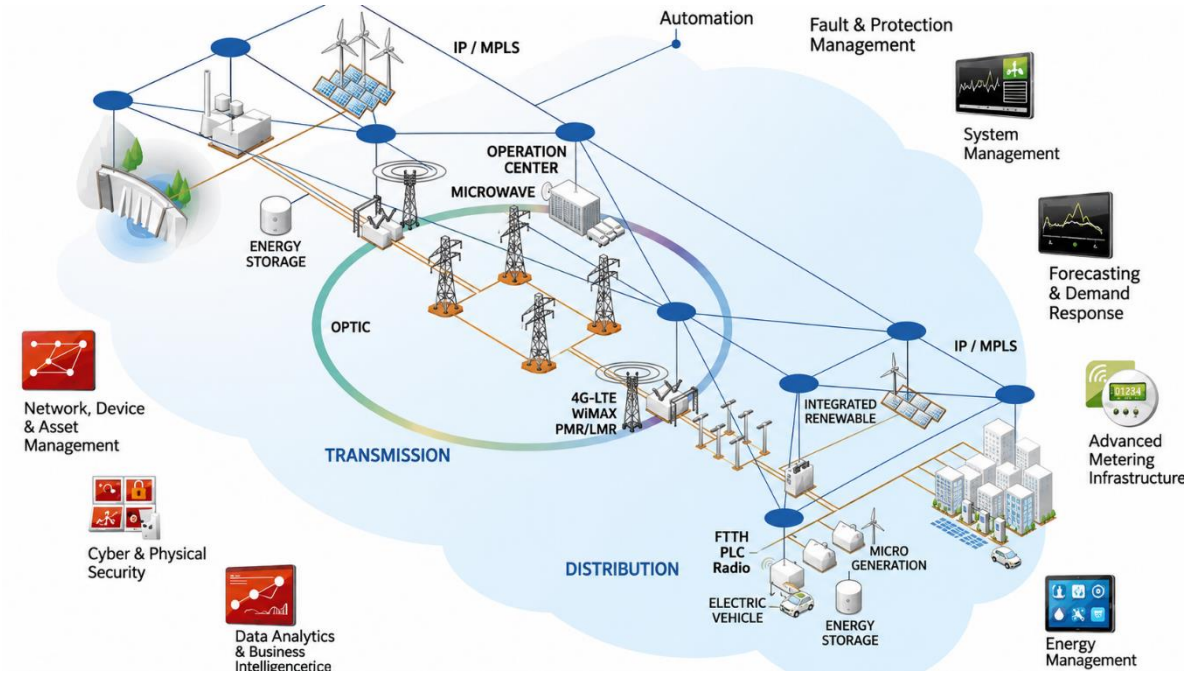


Figure I: Architecture de référence d'une infrastructure critique de type smart grid

I.3. Secteurs des Infrastructures Critiques

Les secteurs d'infrastructures critiques sont catégorisés différemment selon les pays. Les variations entre les nations peuvent s'expliquer par des différences dans les conceptualisations de ce qui est considéré comme critique, ainsi que par des particularités et des traditions propres à chaque pays. Des facteurs sociopolitiques ainsi que des préconditions géographiques et historiques déterminent si un secteur donné est jugé critique. A titre d'exemple, le National Infrastructure Protection Plan (NIPP) aux Etats-Unis divisait les infrastructures critiques en 16 secteurs [2]. De son côté, l'Union européenne a divisé les infrastructures critiques en 13 secteurs [3], l'Australie en 11 secteurs, et l'Inde en 12 [4].

Secteurs	Allemagne	Royaume-Uni	Japon	Etats-Unis	Chine	Inde	Australie
Banque et Finance	■	■	■	■	■	■	■
Gouvernement central / Services gouvernementaux	■	■	■	■		■	
Industrie chimique et nucléaire		■	■	■			
Services d'urgence et de secours		■		■			
Energie	■	■	■	■	■	■	■
Agriculture et alimentation	■	■		■			■
Services de santé	■			■			■
Services d'information / Médias	■		■				
Défense militaire / Armée / Installations de défense		■		■	■		■
Télécommunications / Communications publiques		■		■	■	■	■
Transport	■	■	■	■	■	■	■
Approvisionnement en eau / Assainissement	■	■	■	■	■	■	■
Space		■					■
Industrie manufacturière critique				■			
Technologies de l'information et de la communication (TIC)	■		■	■	■		■
Barrages				■			
Services gouvernementaux électroniques (e-Government)					■		
Entreprises stratégiques et publiques						■	
Enseignement supérieur et recherche							■

Tableau I: Secteurs des infrastructures critiques [1]

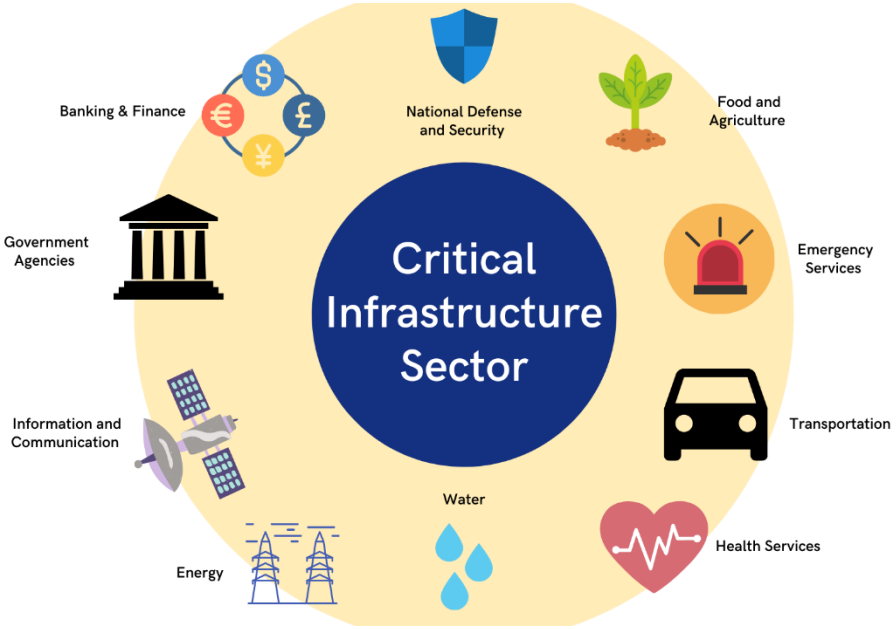


Figure I.1: Secteurs fondamentaux des IC

Le tableau I.1 et la figure I.2 mettent en évidence les secteurs clés qui sont essentiels au monde moderne et qui sont communément reconnus par l'ensemble des nations. Ces secteurs comprennent la finance, l'administration centrale (gouvernement), les télécommunications, les services d'urgence et de secours, l'énergie, les services de santé, l'alimentation, les transports et l'approvisionnement en eau. Ils sont identifiés comme des secteurs fondamentaux en raison de leurs rôles critiques dans la société, où toute perturbation significative pourrait avoir des conséquences désastreuses [1] [5, 6].

I.4 Technologies et normes de communication dans les systèmes à infrastructures critiques

Les infrastructures critiques modernes s'appuient de plus en plus sur les technologies de l'information et de la communication (TIC) pour soutenir leurs activités de surveillance, de contrôle, d'automatisation et d'aide à la décision. Le passage de systèmes industriels isolés à des systèmes cyber-physiques interconnectés a favorisé l'adoption de diverses technologies et normes de communication. Ces technologies permettent un échange de données en temps réel entre capteurs, contrôleurs, sous-stations, centres d'exploitation et plateformes cloud. Par conséquence, la disposition de réseaux de communication fiables, sécurisés et résilients est devenue une exigence fondamentale pour le bon fonctionnement des infrastructures critiques.

I.4.1 Exigences de communication dans les infrastructures critiques

Les systèmes de communication qui desservent les infrastructures critiques doivent satisfaire à des exigences strictes, notamment la fiabilité, la disponibilité, la faible latence, la scalabilité, la sécurité, l'interopérabilité et la résilience [7,8].

Les réseaux doivent garantir un fonctionnement continu même en cas de panne d'équipement, de cyberattaque ou de catastrophe naturelle, tandis que les services de communication doivent rester opérationnels 24 heures sur 24 et 7 jours sur 7 avec un temps d'arrêt minimal.

Les applications telles que la protection des réseaux électriques, l'automatisation industrielle et les interventions d'urgence exigent des délais de communication de seulement quelques millisecondes.

En outre, le réseau doit prendre en charge un nombre croissant de dispositifs, de capteurs et d'utilisateurs (scalabilité), alors que la confidentialité, l'intégrité, l'authentification et la disponibilité des données doivent être protégées contre les cybermenaces .

Différents fournisseurs d'équipements et protocoles de communication doivent fonctionner ensemble de manière transparente, et les systèmes de communication doivent se remettre rapidement des pannes tout en maintenant leur fonctionnement dans des conditions défavorables.

I.4.2 Architecture de communication dans les infrastructures critiques

Les systèmes à infrastructures critiques reposent sur une architecture de communication hiérarchique conçue pour garantir l’échange de données entre de nombreux dispositifs, systèmes de contrôle et plateformes de gestion. Cette architecture facilite la collecte, la transmission, le traitement et l'utilisation des données opérationnelles à travers les différentes couches de l'infrastructure. Bien que la mise en œuvre spécifique puisse varier selon le secteur (énergie, transports, eau, télécommunications) la plupart des infrastructures critiques adoptent une architecture à trois niveaux (figure I.3) composés du niveau terrain, niveau contrôle et du niveau gestion. Chaque niveau remplit des fonctions distinctes et utilise des technologies et protocoles de communication spécifiques pour assurer le fonctionnement global de l'infrastructure [9].

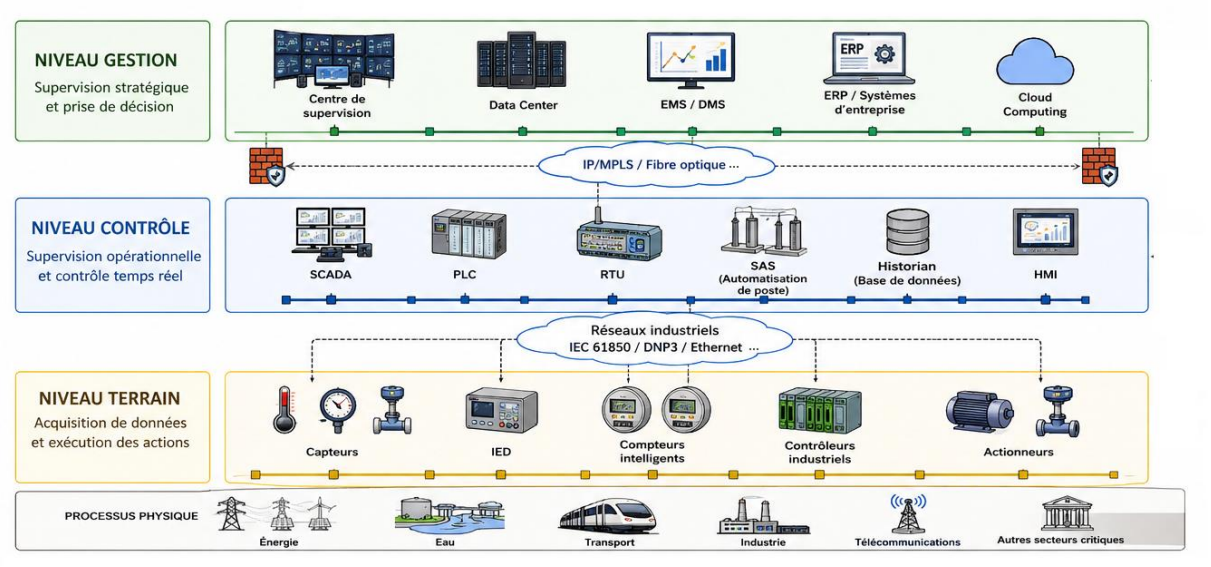


Figure I.2: Niveaux de communications dans les IC

I.4.2.1 Niveau terrain

Le niveau terrain constitue la couche la plus basse de l'architecture de communication et sert d'interface entre l'infrastructure physique et le système de contrôle numérique. Ce niveau est responsable de l'acquisition de données en temps réel à partir des processus physiques et de l'exécution des commandes de contrôle émises par les systèmes de niveau supérieur. Il comprend une grande variété de dispositifs de terrain tels que des capteurs, des actionneurs, des dispositifs électroniques intelligents (IED), des compteurs intelligents et des contrôleurs industriels.

La communication au niveau terrain est généralement assurée par des protocoles et des réseaux de communication industriels offrant fiabilité et performance déterministe. Les réseaux industriels basés sur Ethernet sont devenus de plus en plus populaires en raison de leur large bande passante et de leur interopérabilité. Les technologies de communication série traditionnelles telles que le RS-485 restent largement utilisées en raison de leur simplicité, de leur robustesse et de leur rentabilité dans les environnements industriels. Des protocoles tels que Modbus, PROFIBUS et CAN Bus sont couramment employés pour faciliter la communication entre les dispositifs de terrain [10]. Ces protocoles supportent l'acquisition et le contrôle de données en temps réel tout en fonctionnant dans des conditions environnementales difficiles caractérisées par des interférences électromagnétiques, des variations de température et des vibrations mécaniques.

I.4.2.2 Niveau Contrôle

Le Niveau Contrôle sert de couche intermédiaire entre les dispositifs de terrain et les systèmes de gestion. Sa fonction principale est de collecter les données du niveau terrain, de traiter ces informations, d'exécuter des algorithmes de contrôle et de coordonner le fonctionnement de l'infrastructure. Cette couche fournit des capacités de supervision et de contrôle et garantit que les objectifs opérationnels sont atteints de manière sûre et efficace.

Le Niveau Contrôle comprend généralement des automates programmables industriels (PLC), des unités terminales distantes (RTU), des systèmes d'automatisation de postes électriques (SAS) et des systèmes de supervision, contrôle et acquisition de données (SCADA) [11].

Les technologies de communication utilisées au niveau contrôle incluent généralement l'Ethernet industriel, les réseaux à fibre optique, les systèmes de communication sans fil et des protocoles industriels standardisés tels que DNP3, IEC 60870-5-104 et IEC 61850. Ces technologies garantissent une communication fiable et sécurisée entre les actifs distribués et les installations de contrôle centralisées.

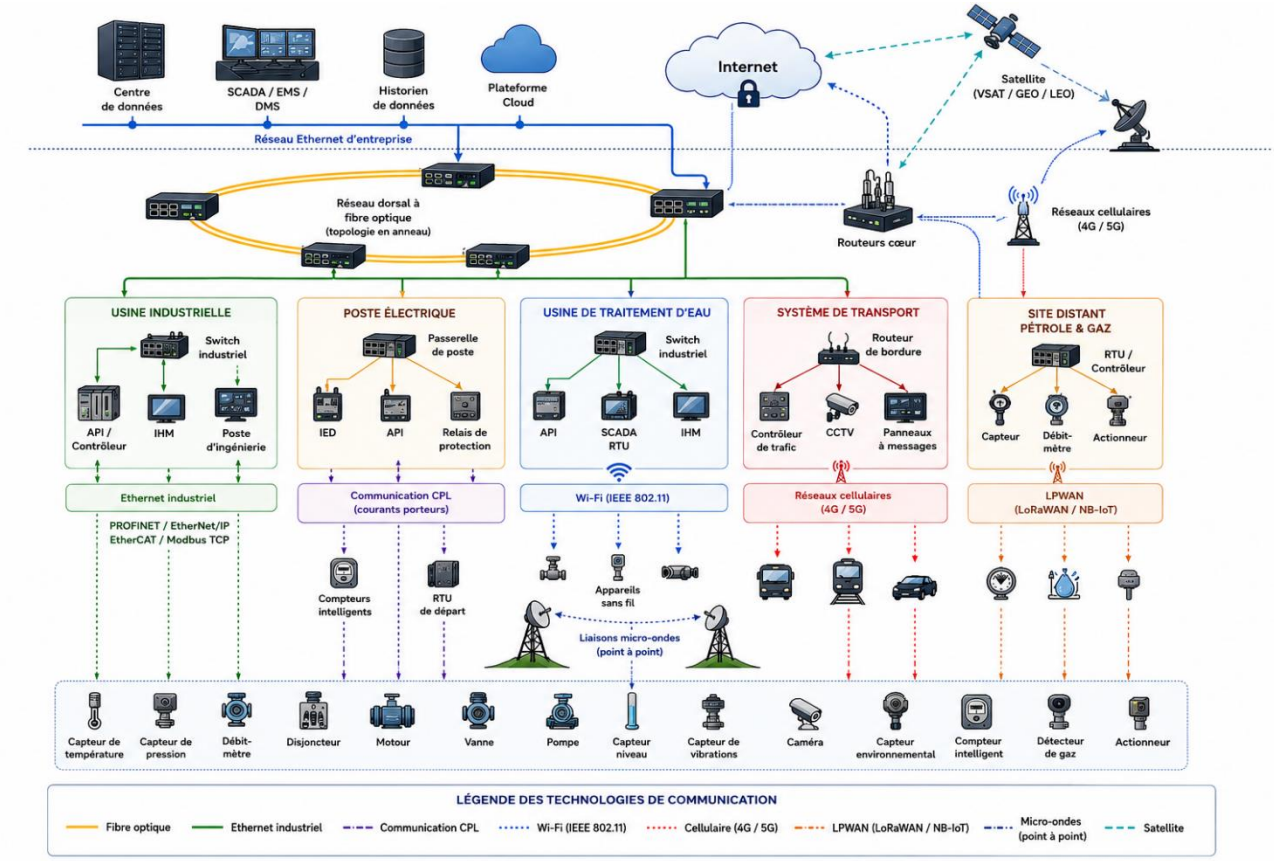
I.4.2.3. Niveau Gestion

Le Niveau Gestion représente la couche la plus élevée de l'architecture de communication et se concentre sur la supervision stratégique, l'optimisation, la planification et la prise de décision à l'échelle de l'entreprise. Alors que les niveaux terrain et contrôle sont principalement concernés par les opérations en temps réel, le niveau gestion transforme les données opérationnelles en informations exploitables qui soutiennent les objectifs de gestion à long terme.

La communication au sein du niveau gestion repose sur des infrastructures de communication à haute capacité et haute fiabilité, incluant les réseaux à fibre optique, les réseaux IP/MPLS, les plateformes de cloud computing et les centres de données sécurisés. Des mécanismes avancés de cybersécurité tels que le chiffrement, l'authentification, les systèmes de détection d'intrusion et la segmentation des réseaux sont essentiels à ce niveau pour protéger les actifs d'information critiques contre les cybermenaces et les accès non autorisés [12].

I.4.3. Technologies de communication dans les infrastructures critiques [1,13]

Dans les infrastructures modernes, de multiples technologies de communication sont combinées pour répondre à divers besoins opérationnels, allant de la communication à courte portée entre capteurs à la connectivité dorsale sur de longues distances.



FigureI.3: Technologies de communication dans les infrastructures critiques

❖ **Réseaux à fibre optique**

Les réseaux à fibre optique représentent l'infrastructure de communication dorsale dans la plupart des systèmes critiques en raison de leurs performances et de leur fiabilité supérieures. Ils utilisent des signaux lumineux transmis à travers des fibres de verre pour transporter de grands volumes de données sur de longues distances avec une perte minimale. Cette technologie est largement déployée les réseaux électriques intelligents, où elle connecte les sous-stations, les centres de contrôle et les ressources énergétiques distribuées., les dorsales de télécommunications et les infrastructures industrielles où une communication à haute capacité et sécurisée est requise.

- **Réseaux Ethernet**

Les réseaux Ethernet industriels sont largement utilisés dans les infrastructures critiques en raison de leur flexibilité, de leur scalabilité et de leurs capacités de communication à haute vitesse. Ils étendent la technologie Ethernet traditionnelle aux environnements industriels en intégrant des performances en temps réel et des fonctionnalités de fiabilité améliorées.

Les normes Ethernet industrielles courantes incluent Ethernet/IP, PROFINET, EtherCAT et Modbus TCP. Ces protocoles permettent une communication transparente entre les dispositifs industriels tels que les contrôleurs, les capteurs et les systèmes de supervision.

- **Communication par courants porteurs en ligne (CPL)**

La communication par courants porteurs en ligne (CPL) est une technologie de communication qui utilise les lignes électriques existantes pour la transmission de données. Cette approche permet de transmettre des signaux de communication sur la même infrastructure que celle utilisée pour la distribution d'énergie électrique, éliminant ainsi le besoin de câblage de communication supplémentaire.

La CPL est couramment utilisée dans les systèmes de comptage intelligent, les applications de surveillance de réseau et les systèmes de gestion de l'énergie domestique. Elle permet aux services publics de collecter à distance les données de consommation et de gérer plus efficacement les réseaux électriques.

- **Wi-Fi (IEEE 802.11)**

Le Wi-Fi, basé sur les normes IEEE 802.11, est largement utilisé dans les environnements de communication locaux au sein des infrastructures critiques. Il fournit une connectivité sans fil pour les dispositifs fonctionnant dans les bâtiments intelligents, les environnements industriels et les systèmes de surveillance locaux.

- **Réseaux cellulaires (4G/5G)**

Les réseaux cellulaires, en particulier les technologies 4G LTE et 5G, jouent un rôle de plus en plus important dans les infrastructures critiques modernes. Ces réseaux permettent une communication sans fil à grande échelle et supportent les applications nécessitant une mobilité et une connectivité à grande échelle.

Dans les infrastructures critiques, les réseaux cellulaires sont utilisés dans les réseaux électriques intelligents, les systèmes de transport intelligents, les systèmes de communication d'urgence et les applications de surveillance à distance. Leur déploiement permet un échange de données en temps réel entre des actifs géographiquement dispersés.

❖ **Technologies LPWAN**

Les technologies de réseau étendu à faible puissance (LPWAN) sont conçues pour la communication longue portée et à faible consommation d'énergie entre dispositifs distribués. Parmi les technologies LPWAN les plus utilisées figurent LoRaWAN et NB-IoT.

LoRaWAN se caractérise par des capacités de communication longue portée, une très faible consommation d'énergie et un faible coût de déploiement. Elle est couramment utilisée dans des applications telles que la surveillance de la qualité de l'eau, la détection environnementale et les déploiements de villes intelligentes où les dispositifs doivent fonctionner pendant de longues périodes sur batterie.

NB-IoT (Narrowband IoT) est une technologie LPWAN basée sur la téléphonie cellulaire, largement utilisée dans la comptabilité des services publics, le suivi d'actifs et la surveillance des infrastructures. Elle fournit une connectivité fiable pour les dispositifs situés dans des environnements difficiles tels que les installations souterraines ou les zones reculées.

❖ **Communications par micro-ondes**

Les systèmes de communication par micro-ondes fournissent des liaisons point à point sans fil souvent utilisées lorsque le déploiement de la fibre optique est peu pratique ou trop coûteux. Ces systèmes transmettent des données à l'aide d'ondes radio haute fréquence entre des antennes placées en visibilité directe.

Les liaisons par micro-ondes sont couramment déployées dans les postes électriques, les infrastructures éloignées et les systèmes de communication militaires. Elles sont particulièrement utiles dans les zones géographiquement difficiles telles que les montagnes, les îles ou les régions rurales.

❖ **Communications par satellite**

Les systèmes de communication par satellite fournissent une couverture mondiale et sont essentiels pour connecter les régions éloignées et inaccessibles au sein des infrastructures critiques. Ces systèmes permettent la communication là où les réseaux terrestres ne sont pas disponibles ou sont peu fiables.

Les technologies satellitaires sont largement utilisées dans les infrastructures maritimes, les installations énergétiques éloignées et les systèmes de communication d'urgence. Elles assurent la connectivité dans les scénarios de reprise après sinistre et supportent les applications de surveillance mondiale.

Les systèmes satellitaires courants incluent le VSAT (Very Small Aperture Terminal), les satellites en orbite géostationnaire (GEO) et les constellations de satellites en orbite basse

(LEO). Les satellites GEO fournissent une couverture stable à long terme, tandis que les systèmes LEO offrent une latence plus faible et des performances de bande passante améliorées.

I.5. Protocoles de communication dans les IC

L'évolution rapide de l'Industrie 4.0, de l'IIoT(Industrial Internet Of Things) et des systèmes cyber-physiques a considérablement accru la diversité des protocoles de communication utilisés dans les infrastructures critiques. Les protocoles traditionnels basés sur la communication série continuent de fonctionner dans les systèmes existants, tandis que les protocoles basés sur Ethernet dominent de plus en plus les nouvelles installations en raison de leur bande passante plus élevée, de leur évolutivité et de leurs capacités d'intégration. Parmi ces protocoles on trouve :

I.5.1 Modbus (RTU et TCP/IP)

Modbus est l'un des protocoles les plus anciens et les plus largement déployés. Développé à l'origine par Modicon en 1979, il reste un standard de fait en raison de sa simplicité, de son ouverture et de son interopérabilité. Modbus RTU fonctionne sur des liaisons de communication série telles que RS-485, tandis que Modbus TCP/IP étend le protocole aux réseaux Ethernet. Modbus utilise une architecture maître-esclave, où un seul équipement maître peut communiquer avec jusqu'à 247 équipements esclaves, permettant une surveillance et un contrôle centralisés des processus industriels. Les vitesses de communication typiques vont de 9,6 kbit/s à 115,2 kbit/s pour Modbus RTU et de 10 Mbit/s à 100 Mbit/s pour Modbus TCP/IP, selon l'infrastructure réseau. Son faible coût d'implémentation et le large support des fabricants le rendent très adapté à la communication PLC-IHM, aux systèmes de gestion de l'énergie, aux capteurs industriels et aux applications SCADA. Cependant, Modbus natif ne dispose pas de mécanismes de cybersécurité intégrés tels que l'authentification et le chiffrement [14].

I.5.2. PROFINET

PROFINET (Process Field Network) est un protocole Ethernet industriel de haute performance développé par Siemens et l'organisation PROFIBUS & PROFINET International (PI). Conçu pour l'automatisation d'usine et le contrôle industriel en temps réel, il supporte la communication déterministe, des temps de cycle rapides, le diagnostic des équipements et une intégration transparente avec l'infrastructure Ethernet standard [15].

Grâce à ses classes de performance Temps Réel RT (Real Time) et Temps Réel Isochrone IRT (Isochron Real Time), PROFINET couvre un large éventail d'applications, de la surveillance de processus au contrôle de mouvement à grande vitesse. Il fonctionne à des vitesses de 100 Mbit/s à 1 Gbit/s, assurant un échange rapide de données entre contrôleurs, équipements de terrain et systèmes de supervision, tout en restant compatible avec les réseaux TCP/IP conventionnels.

PROFINET simplifie l'intégration et la maintenance grâce à l'échange à chaud des équipements sans perturbation majeure du réseau, à l'adressage automatique qui facilite le déploiement et le remplacement, aux serveurs web intégrés permettant l'accès aux paramètres, états et diagnostics, ainsi qu'aux fichiers GSDML standardisés qui assurent l'interopérabilité entre fabricants.

Ces caractéristiques font de PROFINET un protocole de référence dans les systèmes d'automatisation modernes. Il est largement utilisé dans les réseaux d'automates Siemens S7-1200 et S7-1500, les systèmes de fabrication à grande vitesse, les lignes de production robotisées et les applications de contrôle de mouvement de haute précision. Il est également adopté dans les environnements nécessitant une intégration étroite entre technologies opérationnelles OT (Operational Technology) et informatiques IT (Information Technology), supportant ainsi l'Industrie 4.0 et la fabrication intelligente.

I.5.3. EtherNet/IP

EtherNet/IP (Ethernet Industrial Protocol) est un protocole Ethernet industriel basé sur le Common Industrial Protocol (CIP) [16]. Il utilise les technologies Ethernet et TCP/IP standard tout en fournissant des services spécifiques pour le contrôle et la surveillance en temps réel, permettant la communication entre automates (PLC), dispositifs d'E/S répartis, variateurs, capteurs et systèmes SCADA.

Construit sur le même CIP que DeviceNet, il assure la cohérence entre différentes plates-formes de communication. Il supporte des vitesses de 10 Mbit/s à 1 Gbit/s et offre transfert de données à grande vitesse, évolutivité, large support des fournisseurs et intégration avec les réseaux d'entreprise existants. Son utilisation d'Ethernet standard non modifié réduit les coûts de déploiement et simplifie l'intégration avec les infrastructures réseau en place.

EtherNet/IP utilise un modèle Producteur-Consommateur : un seul paquet de données peut être consommé simultanément par plusieurs équipements via multicast, ce qui réduit le trafic réseau et améliore l'évolutivité, particulièrement dans les grands systèmes

d'automatisation. La configuration est facilitée par des fichiers EDS (Electronic Data Sheet) standardisés, qui décrivent les capacités, paramètres et exigences des équipements, simplifiant ainsi la mise en service et l'interopérabilité entre fournisseurs.

EtherNet/IP est particulièrement répandu dans les installations industrielles nord-américaines et largement utilisé dans les automates Allen-Bradley ControlLogix et CompactLogix, les usines automobiles, les industries agroalimentaires, les lignes d'emballage et les systèmes de manutention. Sa flexibilité, sa haute vitesse et ses capacités d'intégration IT en font l'un des protocoles Ethernet industriels les plus adoptés dans la fabrication et l'automatisation des processus modernes.

I.5.4. PROFIBUS

PROFIBUS (Process Field Bus) est un protocole de terrain éprouvé largement utilisé dans l'automatisation des processus et des usines. Avant l'émergence des technologies Ethernet industrielles, il servait de standard de communication dominant dans de nombreux secteurs industriels [17]. Bien que de nombreuses nouvelles installations migrent vers PROFINET, PROFIBUS reste largement déployé dans les environnements existants et les infrastructures critiques où la durée de vie des équipements dépasse souvent vingt ans.

Il fournit une communication déterministe et supporte les systèmes de contrôle distribués grâce à une connectivité robuste au niveau terrain. PROFIBUS fonctionne à des vitesses de 9,6 kbit/s à 12 Mbit/s, le rendant adapté aussi bien à la surveillance de processus à basse vitesse qu'à des tâches d'automatisation plus rapides. Il utilise des supports physiques tels que RS-485 pour la communication standard et des fibres optiques pour les environnements nécessitant longue distance ou haute immunité aux interférences électromagnétiques, permettant un déploiement dans des conditions industrielles difficiles.

Le protocole supporte à la fois l'échange de données cyclique (contrôle en temps réel et mises à jour périodiques) et la communication acyclique (paramétrisation, diagnostic, configuration). Il offre également les modes maître-esclave et pair-à-pair, autorisant des architectures de contrôle centralisées ou distribuées selon les besoins. L'intégration des équipements est standardisée via des fichiers GSD (Generic Station Description), qui définissent les capacités et paramètres des équipements, simplifiant ainsi la configuration et l'interopérabilité.

En raison de sa robustesse et de son adoption de longue date, PROFIBUS reste largement utilisé dans les réseaux d'automates Siemens S7, les industries de process (chimie,

pétrole et gaz, produits pharmaceutiques), les systèmes d'automatisation d'usine avec architectures d'E/S distribuées, ainsi que dans le contrôle de mouvement et les systèmes d'entraînement où une communication fiable et déterministe est essentielle. Malgré la transition vers PROFINET, il continue de jouer un rôle significatif dans la maintenance et l'exploitation des infrastructures industrielles existantes.

I.5.5. EtherCAT

EtherCAT (Ethernet for Control Automation Technology) est un protocole Ethernet industriel à haute vitesse spécifiquement optimisé pour le contrôle déterministe en temps réel [18]. Idéal pour la robotique, le contrôle de mouvement et les systèmes d'automatisation haute performance, il atteint des temps de cycle extrêmement courts (environ 100 microsecondes) et une synchronisation précise au niveau de la microseconde.

Contrairement aux méthodes Ethernet traditionnelles qui reposent sur un traitement de type store-and-forward, EtherCAT traite les données à la volée lorsque la trame traverse chaque nœud du réseau. Cette architecture réduit significativement la latence et assure un comportement hautement déterministe, essentiel pour les systèmes de contrôle industriel de précision.

EtherCAT supporte des configurations réseau très évolutives, permettant jusqu'à 65 535 équipements sur un seul segment de communication tout en maintenant une gigue faible et une haute précision temporelle, garantissant un fonctionnement étroitement coordonné. De plus, il s'agit d'une technologie gratuite et ouverte (sans frais de licence), ce qui favorise son adoption généralisée dans la recherche académique et les applications industrielles.

Grâce à ces capacités, EtherCAT est largement utilisé dans les systèmes de contrôle de mouvement de haute précision (robotique, machines CNC), les équipements de fabrication de semi-conducteurs, les systèmes d'emballage nécessitant une synchronisation serrée, ainsi que les plates-formes avancées de test et de mesure. Son ultra-faible latence, son évolutivité et sa synchronisation précise en font l'un des principaux protocoles pour l'automatisation de nouvelle génération.

I.5.6. CANopen et DeviceNet

CANopen et DeviceNet sont des protocoles de communication industrielle construits sur le standard Controller Area Network (CAN). Ils partagent la même couche physique et de liaison de données (CAN) mais diffèrent par leur couche application et leurs domaines cibles : CANopen met l'accent sur l'interopérabilité et les performances en temps réel, tandis que DeviceNet étend CAN pour le contrôle industriel avec des profils d'équipement standardisés et des fonctions de gestion de réseau.

CANopen est largement utilisé en Europe pour le contrôle de mouvement, les systèmes embarqués et les dispositifs médicaux. Il fournit un cadre flexible et standardisé pour la communication et la configuration dans les systèmes d'automatisation distribués. Il fonctionne à des vitesses allant jusqu'à 1 Mbit/s et supporte jusqu'à 127 nœuds par réseau. Grâce à son dictionnaire d'objets (OD) qui structure tous les paramètres de communication et d'application, et aux fichiers standardisés EDS et DCF facilitant la configuration et l'interopérabilité, CANopen assure une intégration cohérente entre fabricants. Sa fiabilité et son comportement déterministe le rendent particulièrement adapté au contrôle de mouvement (servovariateurs, moteurs pas à pas), aux dispositifs médicaux, à l'automatisation de laboratoire, aux systèmes embarqués et aux bancs d'essai automobiles [19].

DeviceNet est couramment utilisé pour connecter capteurs, actionneurs, démarreurs et variateurs aux automates (PLC), notamment dans les environnements Rockwell Automation. Il étend CAN avec des services additionnels de couche application pour l'automatisation industrielle. Il supporte trois vitesses (125, 250 et 500 kbit/s selon la taille du réseau et la longueur des câbles) et jusqu'à 64 nœuds par réseau. Sa caractéristique distinctive est sa capacité à transporter alimentation et signaux de communication sur le même câble (configurations fin ou épais). Il utilise les modèles producteur-consommateur et maître-esclave pour optimiser l'échange de données, et la configuration se fait via des fichiers EDS standardisés [20].

I.5.7. MQTT

MQTT (Message Queuing Telemetry Transport) est un protocole de messagerie léger de type publication-abonnement, de plus en plus utilisé dans l'Internet industriel des objets (IIoT) et les systèmes industriels connectés au cloud. Contrairement aux protocoles traditionnels basés sur le polling, il permet une communication pilotée par les événements grâce à une architecture basée sur un courtier (broker), réduisant ainsi la consommation de bande

passante et améliorant l'évolutivité. Sa simplicité, son efficacité et sa faible charge ont conduit à son adoption généralisée dans les infrastructures industrielles modernes [21].

MQTT repose sur un modèle publication-abonnement où tous les messages sont acheminés via un courtier (broker) qui gère la communication entre les clients, assurant le découplage des systèmes et supportant l'échange de données de plusieurs à plusieurs sans connexion directe entre émetteur et récepteur, ce qui rend ce protocole très efficace pour les systèmes impliquant un grand nombre de capteurs et d'équipements connectés. Sa légèreté et sa flexibilité lui permettent de fonctionner avec des ressources minimales, idéal pour les environnements contraints (équipements alimentés par batterie, réseaux à faible bande passante), et il opère sur TCP/IP standard, WebSockets ou réseaux cellulaires, autorisant un déploiement flexible local ou distant. MQTT définit trois niveaux de qualité de service (QoS) QoS 0 (au plus une fois), QoS 1 (au moins une fois) et QoS 2 (exactement une fois) – permettant d'équilibrer fiabilité et charge de communication selon les besoins. Enfin, il supporte des fonctionnalités avancées telles que les messages retenus (stockage de la dernière bonne valeur pour les nouveaux abonnés) et le mécanisme Last Will and Testament (LWT) (notification de défaut en cas de déconnexion inattendue), améliorant ainsi la fiabilité et la connaissance des défauts dans les applications critiques.

MQTT connaît un déploiement étendu dans les systèmes de collecte de données IIoT, les plates-formes cloud (AWS IoT Core, Microsoft Azure IoT Hub, Google Cloud IoT), la surveillance à distance d'équipements de terrain sur réseaux cellulaires, les applications mobiles pour la surveillance d'usine, les pipelines de données edge-to-cloud et les réseaux de capteurs sans fil alimentés par batterie. Il constitue ainsi un protocole fondamental pour la transformation numérique industrielle moderne.

I.5. Vulnérabilités et menaces dans les IC

Les infrastructures critiques, comme mentionné précédemment, jouent un rôle crucial dans le fonctionnement de la société. C'est pourquoi les pays cherchent à améliorer leur efficacité tout en réduisant les délais et les coûts de production. Cependant, certaines des améliorations que les pays mettent en œuvre pour atteindre cet objectif ont créé des vulnérabilités dans les systèmes d'infrastructures critiques. Ces vulnérabilités permettent alors aux attaquants de contourner la sécurité de ces systèmes et d'obtenir des accès avec des privilèges qui ne devraient pas leur être accordés.

Selon le CSIS (Center forStrategic and International Studies), les infrastructures critiques deviennent plus vulnérables, notamment dans les secteurs de la défense , des services publics, ou encore de la communication. Elles représentent une part importante des cyberincidents enregistrés depuis 2006 [22]. La figure I.5 présente les attaques significatives recensées par infrastructure critique.

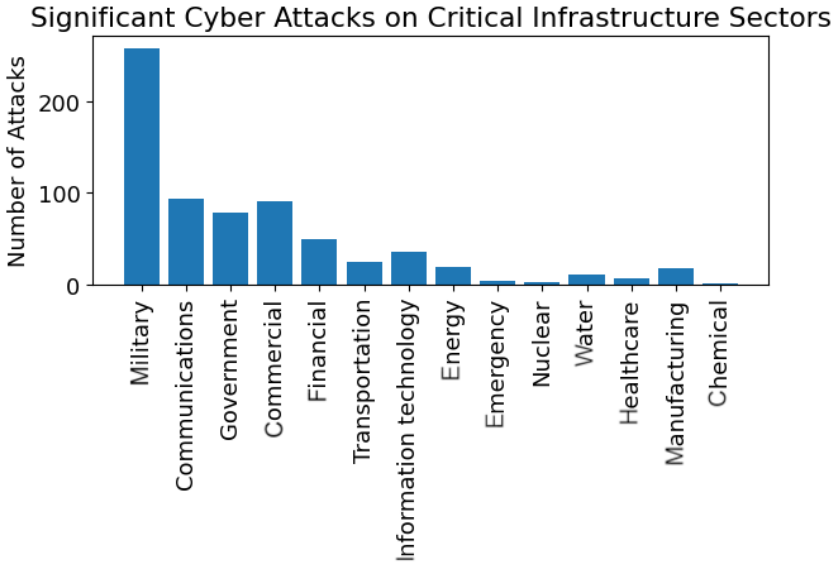


Figure I.4: Cyberattaques significatives par secteur d'infrastructure critique depuis 2006

Des vulnérabilités récentes (CVE-2023-20080 sur Cisco IOS, CVE-2023-28231 sur DHCPv6 Windows, CVE-2024-38063 sur TCP/IP IPv6 Windows) ont montré que des protocoles réseau matures et largement déployés contiennent des failles exploitables menaçant la disponibilité des infrastructures critiques. La NSA a publié des recommandations sur la sécurité IPv6 pour les opérateurs d'infrastructures critiques. Au-delà des vulnérabilités logicielles, des défis systémiques persistent : les infrastructures critiques sont interconnectées, créant des chaînes de vulnérabilités en cascade. L'hétérogénéité des systèmes (équipements anciens sans sécurité intégrée aux côtés d'objets IoT modernes) complique l'application de politiques de sécurité unifiées. De plus, les contraintes opérationnelles (haute disponibilité, latence stricte) empêchent souvent de mettre hors ligne les équipements pour les mettre à jour [23].

D’autre part, les menaces proviennent d'acteurs variés : cybercriminels opportunistes jusqu'aux adversaires sophistiqués. Les attaques DDoS persistent contre les infrastructures connectées. Au-delà du déni de service, les adversaires ciblent l'intégrité et la confidentialité des systèmes de contrôle industriels. Les objets IoT déployés dans les environnements critiques souffrent fréquemment d'une authentification inadéquate, de communications non chiffrées et

d'une validation insuffisante des entrées, offrant des points d'entrée aux attaquants. Le tableau I.2 [23] récapitule les principales catégories de vulnérabilités rencontrées dans les systèmes à infrastructures critiques.

Catégorie de vulnérabilité	**Description**	**Exemples de manifestations**	**Impact potentiel**
Vulnérabilités des protocoles	Failles dans les protocoles réseau fondamentaux	CVE-2023-20080 (DHCPv6), CVE-2024-38063 (IPv6 TCP/IP)	Exécution de code à distance, déni de service, contournement de la segmentation réseau
Sécurité des objets IoT	Authentification faible, communications non chiffrées, mises à jour inadéquates	Identifiants par défaut, protocoles en clair, micrologiciel obsolète	Compromission de l'appareil, recrutement dans un botnet, reconnaissance du réseau
Intégration de systèmes hérités (legacy)	Systèmes conçus sans considérations de sécurité interfacant avec des réseaux modernes	Protocoles SCADA non chiffrés, contrôles d'accès inadéquats	Perturbation opérationnelle, exfiltration de données, impacts sur la sécurité physique
Risques liés à la chaîne d'approvisionnement	Composants ou logiciels compromis dans les systèmes d'infrastructure	Micrologiciel malveillant, matériel piégé	Accès persistant à long terme, compromission généralisée
Faiblesses de configuration	Paramètres de sécurité incorrects et durcissement inadéquat	Interfaces de gestion ouvertes, permissions excessives	Accès non autorisé, escalade de privilèges

Tableau I.1: Principales catégories de vulnérabilités dans les IC

I.6. Types de cyber-attaques dans les IC

Les infrastructures critiques font l'objet d'un nombre considérable de cyberattaques. Dans ce qui suit, nous décrivons succinctement les principales techniques d'attaque employées.

1.6.1. Attaque par Ransomware

Un ransomware a pour action malveillante de chiffrer, verrouiller ou exfiltrer des données.La diversité des systèmes d'exploitation implique que des bibliothèques et fonctions spécifiques au système sont utilisées par le ransomware pour mener ses actions malveillantes. Ils ciblent principalement les PC/postes de travail sous Windows . Dans l'économie de la cybercriminalité, certains groupes opèrent sous la forme de Rançongiciel en tant que société RAAC (Ransomware-as-a-Corporation). Les attaquants RAAC publient fréquemment des communiqués de presse et utilisent un langage d'entreprise dans leurs communications. Si la rançon n'est pas payée, les systèmes opérationnels des victimes restent inaccessibles et toute information personnelle critique exfiltrée est publiée sur un site de fuite du dark web pour nuire à la réputation et aux processus commerciaux de l'entreprise [24].

L'installation d'appareils électroniques intelligents dans les IC rendait celles-ci ainsi que leurs systèmes cyber-physiques plus susceptibles d'être ciblées par les Ransomware. Alors que les technologies intelligentes continuent de se déployer et de s'intégrer dans les foyers, les transports, les bâtiments et les villes, elles deviendront une cible croissante pour l'évolution future des Ransomware ciblant ce nouvel environnement. Ainsi, les Ransomware ciblant les appareils électroniques intelligents industriels deviendront plus répandus.

Le plus souvent, les e-mails sont le vecteur de distribution des Ransomware. Les e-mails malveillants contiennent le Ransomware en pièce jointe. Ces messages sont souvent envoyés sous forme de spam à un maximum d'adresses e-mails ou peuvent être dirigés et adaptés à des individus ou organisations spécifiques. La pièce jointe peut fournir un lien ou un fichier qui déclenche l'installation du Ransomware [25].

1.6.2. Déni de service (DoS/DDoS)

Dans une attaque par déni de service (DoS :Denial of Service) et déni de service distribuée (DDOS :Distributed Denial of Service), l'attaquant empêche l'utilisateur légitime d'accéder à une ressource. L'attaquant peut réduire l'accès de l'utilisateur au serveur en inondant le réseau, c'est-à-dire en augmentant le trafic pour perturber l'accès à un service. L'attaquant tente également de rompre la connectivité entre deux systèmes. Les attaques par saturation (flooding) font que le système reçoit un trafic trop important pour que les serveurs puissent le gérer. Saturer un système le ralentit et peut finir par l'arrêter complètement.

L'installation croissante d'appareils électroniques intelligents dans les systèmes IC et les appareils de l'IoT, tels que les maisons connectées et les appareils électroménagers intelligents, accroît également les dommages potentiels aux infrastructures critiques causés par les attaques DoS [26].

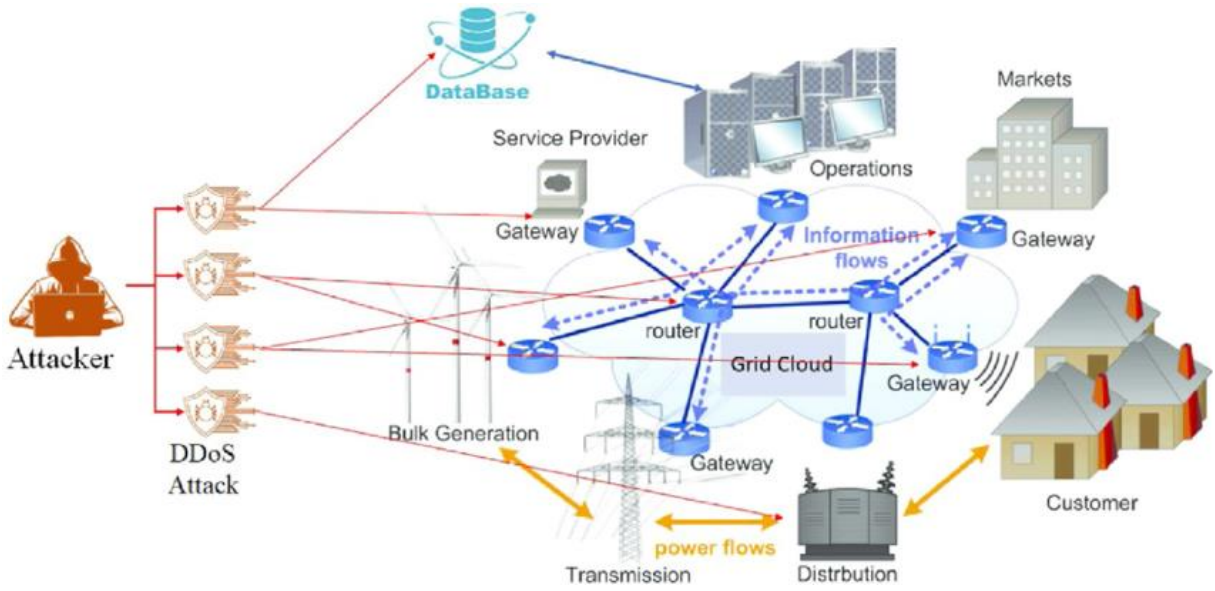


Figure I.5: Attaque DDOS contre une IC (exemple de smart grid)

1.6.3. Attaque de l'homme du milieu (Man-in-the-Middle)

L'attaque de l'homme du milieu (MITM) est un type de cyberattaque où un tiers s'insère entre deux nœuds de communication et tente de rester indétecté. L'attaquant MITM peut modifier les informations routées avant qu'elles n'atteignent l'autre nœud. Cette cyberattaque accède, lit, modifie ou altère les informations secrètes sans que la victime détecte la manipulation. Une capacité consiste à injecter de nouveaux messages, une autre à intercepter tous les messages. Malgré le chiffrement, un attaquant MITM compromet avec succès les échanges entre deux systèmes. Le MITM est soit un auditeur passif, soit il imite l'une des parties et manipule les données envoyées. Les objectifs d'une telle attaque peuvent être multiples : utiliser les données interceptées pour une action ultérieure ou modifier les données avant qu'elles n'atteignent l'autre partie. Les techniques d'exploitation utilisées pour mener une attaque MITM incluent l'usurpation IP (IP spoofing), l'usurpation ARP (ARP spoofing), l'usurpation GNSS (système mondial de navigation par satellite) et l'usurpation DNS (DNS spoofing) [27].

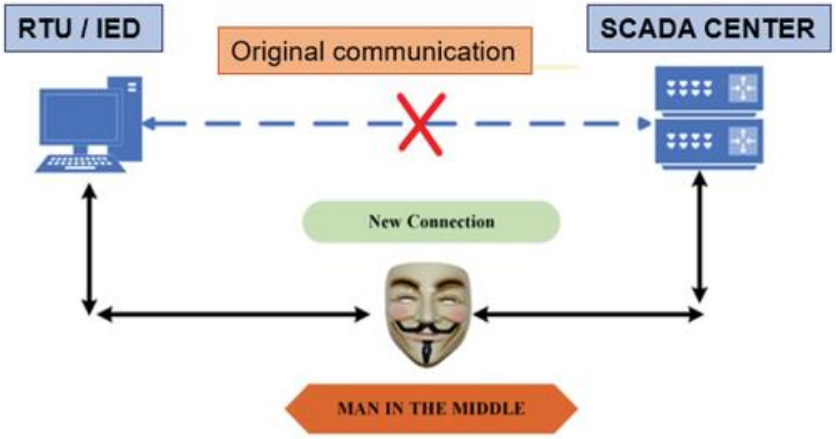


Figure I.6: Principe général du l'attaque MIMT

1.6.3.1. Usurpation IP (IP Spoofing)

Dans l'usurpation IP, l'attaquant modifie l'adresse source dans l'en-tête du paquet IP pour faire croire au destinataire que le paquet provient d'un site de confiance. Du côté de la victime, les paquets seront reçus comme s'ils provenaient d'une source fiable. Cependant, la source IP indiquée dans le paquet est modifiée et ne représente pas la source réelle [28].

1.6.3.2. Usurpation ARP (ARP Spoofing)

L'usurpation ARP consiste à envoyer un faux message de réponse ARP à la passerelle réseau par défaut, prétendant associer l'adresse MAC à l'adresse IP de la cible. Ce protocole ARP traduit les adresses IP en adresses MAC. Les paquets ARP MITM sont transmis sur le réseau local (LAN) en envoyant des paquets ARP malveillants à une passerelle par défaut sur ce réseau [29]. La requête de réassociation de l'attaquant peut lui permettre d'apparaître comme la passerelle par défaut pour le trafic ; ainsi, tous les autres hôtes du réseau transmettront leurs données via l'attaquant MITM.

1.6.3.3. Usurpation DNS (DNS Spoofing)

Dans l'usurpation DNS, l'adresse IP d'un enregistrement DNS est remplacée par une adresse IP contrôlée par l'attaquant. Cela redirige le trafic Internet vers des sites web frauduleux ressemblant aux destinations prévues [30].

1.6.3.4. Détournement HTTPS/SSL (Hijacking)

Les données volées peuvent être déchiffrées par plusieurs méthodes, notamment l'usurpation HTTPS, le détournement SSL, le dépouillement SSL (SSL stripping), etc. Dans

l'usurpation HTTPS, l'attaquant utilise un domaine qui semble identique à celui du site web cible. Dans le détournement SSL, l'attaquant transmet les clés d'authentification produites à la fois au client et à l'application lors d'une prise de contact TCP (handshake) [30]. Cela semble, à tous égards, être une connexion sécurisée alors que l'attaquant MITM contrôle toute la session. Dans l'attaque SSL stripping, l'attaquant envoie une version décodée du site de l'application au client tout en maintenant la session établie avec l'application. Pendant ce temps, toute la session du client est visible par l'attaquant.

1.6.4. Hameçonnage (Phishing) et attaques à distance

Les attaques par hameçonnage et les attaques à distance reposent sur des méthodes d'ingénierie sociale conçues pour amener la victime à révéler des informations sensibles ou à utiliser un logiciel malveillant. L'hameçonnage est très répandu dans les cyberattaques contre les infrastructures critiques IC et il est identifié dans de nombreuses cyberattaques majeures (tableauI.3). Les attaquants envoient des communications frauduleuses pour contraindre une victime à partager des identifiants classifiés ou d'autres informations. Les identifiants obtenus peuvent être utilisés pour mener d'autres attaques, telles que l'installation de logiciels malveillants, l'accès à distance ou le vol d'informations. Les attaquants peuvent également exiger une rançon en menaçant de publier les informations. Les méthodes d'hameçonnage sont également utilisées pour introduire des infections par ransomware dans l'infrastructure réseau de la victime [31].

Le déroulement de la cyberattaque par hameçonnage et exécution à distance contre l'infrastructure critique du système de distribution d'énergie est illustré à la figure I.8. Le processus commence par une phase de reconnaissance, suivie d'une campagne de hameçonnage, de l'obtention d'un accès, de la création d'un tunnel vers l'OT, de l'installation d'un logiciel malveillant dans l'OT, et enfin de l'utilisation d'interfaces homme-machine pour saboter les systèmes physiques sur le terrain.

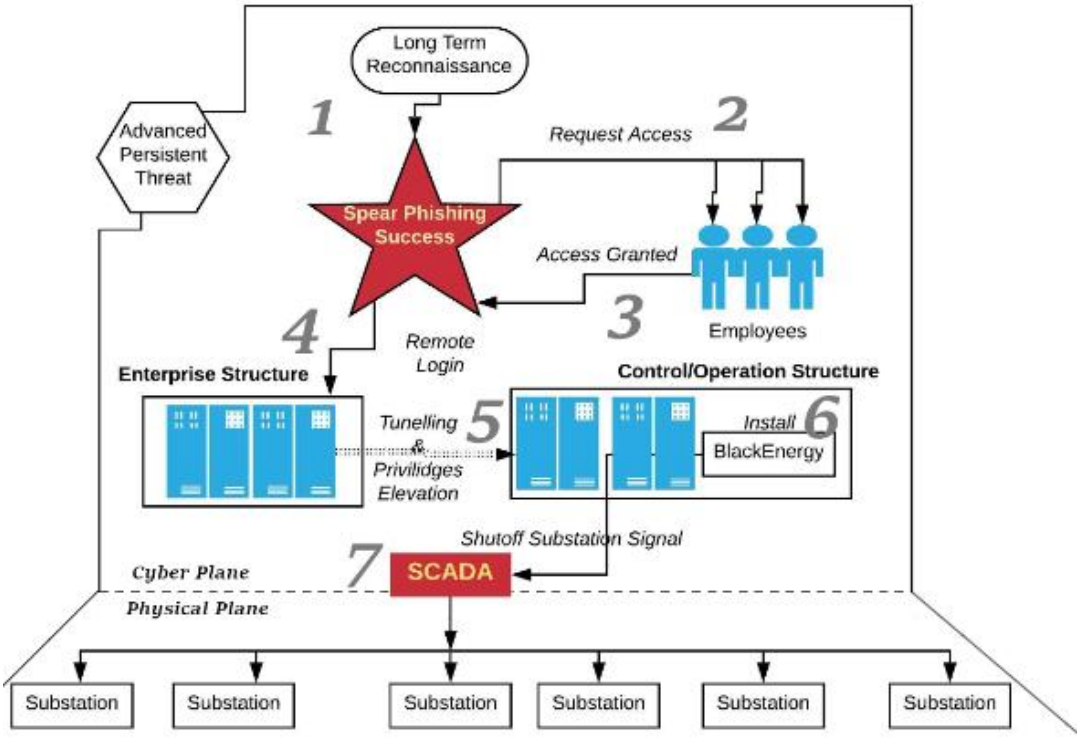


Figure I.7: Processus cyberattaque par hameçonnage et exécution à distance

1.6.5. Attaque par injection de fausses données (FDIA)

L'attaque par injection de fausses données (False Data Injection Attack, FDIA) est une cyberattaque qui vise à corrompre les données de contrôle d'un système à IC, par exemple dans un réseau électrique ou un système industriel. En corrompant les mesures ou les commandes (injection de paramètres ou de commandes), l'attaquant peut tromper les opérateurs ou déstabiliser le système physique. La FDIA se divise en trois types principaux [32] :

- **FDIA ciblée avec contraintes :** Dans ce type d'attaque, les données sont injectées après une analyse minutieuse du système cible. L'attaquant insère une quantité connue de données falsifiées, mais celles-ci restent cohérentes avec le comportement attendu du système afin de ne pas éveiller les soupçons. L'objectif est de passer inaperçu tout en influençant discrètement le processus.
- **FDIA ciblée sans contraintes :** Ici, l'attaquant cherche à corrompre directement les valeurs de certaines variables critiques. Ces variables modifiées entraînent à leur tour la corruption d'autres variables dépendantes, ce qui peut provoquer des dysfonctionnements en cascade dans le système. L'attaque est plus agressive et plus facilement détectable, mais ses effets peuvent être très étendus.

- **FDIA aléatoire :** Dans ce type d'attaque, l'attaquant injecte des paquets de données falsifiées de manière aléatoire, sans tenir compte des valeurs réelles du système. Il n'y a aucune tentative de rendre les données injectées plausibles ou cohérentes. Cette approche a pour but de créer du bruit, de saturer le système d'informations erronées ou de provoquer des alertes intempestives, masquant éventuellement une autre attaque plus ciblée.

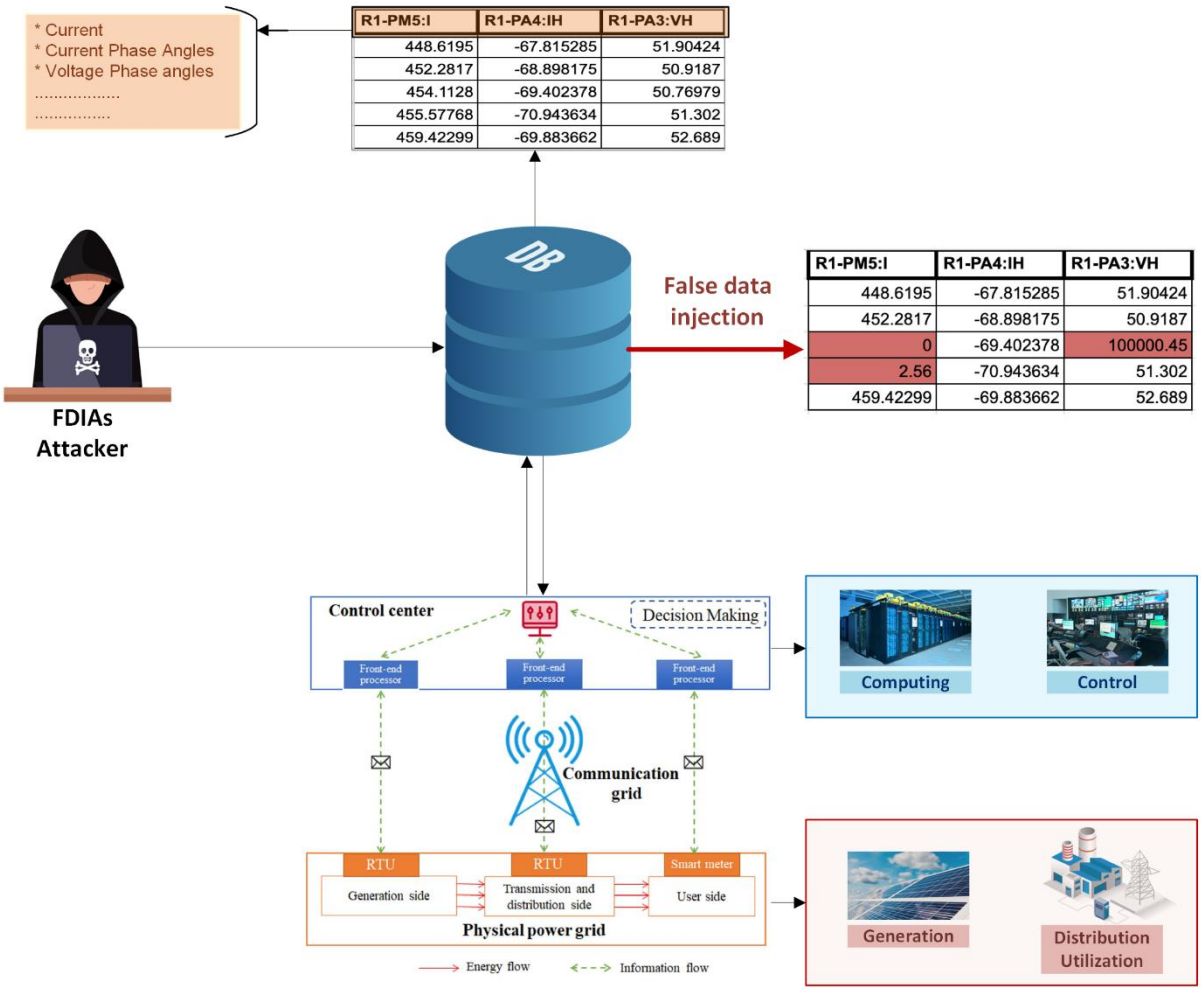


Figure I.8: Attaque FDIA

1.6.6. Attaque par les vers

Un ver informatique est un virus informatique caractérisé comme un logiciel malveillant auto-réplicatif qui se propage sur les réseaux en exécutant des charges utiles disruptives [33]. Un ver cible des hôtes selon les types de balayage suivants :

- **Un balayage actif sélectif aléatoire ou séquentiel :** dans ce type, le ver recherche les hôtes vulnérables.
- **Un balayage par liste de cibles :** le ver crée une liste de cibles, puis recherche les hôtes susceptibles d'être compromis.
- **Un balayage routable :** celui-ci utilise des informations sur un réseau pour sélectionner et balayer l'espace d'adresses IP.

L'utilisation d'une adresse IP routable permet au ver de se propager rapidement et efficacement, tout en échappant à certaines méthodes de détection. Une autre caractéristique d'un ver est l'espace ou le support cible par lequel il se propage. Cela inclut Internet, les courriels, les réseaux P2P, les clés USB locales, etc. Le ver se propage soit de manière autonome, soit via un second canal. Dans la méthode du second canal, la charge utile principale du logiciel malveillant est téléchargée à distance par l'installeur de base. L'activation d'un ver sur un système exploite une vulnérabilité de l'hôte, et le ver peut se protéger en modifiant son code binaire par chiffrement.

1.6.7. Exemples de cyber attaques contre les IC

Des événements d'attaque importants sont présentés dans le tableau I.3 [34], chacun d'entre eux ayant causé des dommages considérables. Ces incidents majeurs, qui incluent des ransomwares, des attaques DOS, attaques FDIA, Physhing… , illustrent la diversité des cybermenaces auxquelles les infrastructures critiques sont confrontées.

Année	Cible	Type d'attaque	Impact
2010	Centrifugeuses nucléaires Iraniennes (Natanz)	Ver (Stuxnet)	Destruction physique des centrifugeuses
2015	Réseau électrique ukrainien	Hameçonnage, accès distant	Panne affectant ~225 000 clients
2016	Station de transmission de Kiev (Ukraine)	Hameçonnage, sabotage distant	Panne d'une heure
2017	Infrastructure mondiale (transport, santé, etc.)	Ransomware (WannaCry, NotPetya)	Pertes financières, perturbations généralisées
2019	Systèmes industriels (multi-secteurs)	Injection de fausses données (FDIA)	Corruptions de données de contrôle
2020	Chaîne d'approvisionnement logicielle (SolarWinds)	Compromission de mise à jour	Accès non autorisé à des milliers d'organisations
2021	Oléoduc colonial (Etats-Unis)	Ransomware (DarkSide)	Pénurie de carburant, paiement d'une rançon

Tableau I.2: Cyberattaques significatives des dernières années

1.7. Conclusion

Ce chapitre a passé en revue les systèmes à infrastructures critiques IC. Après avoir défini ces dernières et identifié les secteurs stratégiques qu'elles couvrent (énergie, eau, transport, santé, etc.), nous avons examiné les technologies de communication et protocoles industriels qui les sous-tendent. L'analyse des vulnérabilités inhérentes à ces systèmes a révélé une surface d'attaque étendue et particulièrement exposée. Nous avons ensuite détaillé les principaux types de cyberattaques exploitant ces failles : ransomware, DoS/DDoS, MITM, FDIA, vers et le hameçonnage, vecteur d'entrée privilégié. Chacune de ces menaces a démontré sa capacité à provoquer des dommages physiques, financiers et réputationnels majeurs.

Face à ce constat, il apparaît clairement que la sécurité des infrastructures critiques n'est plus une option mais une nécessité absolue. Mieux s'en protéger et atteindre un niveau de sécurité approprié devient alors une tâche primordiale, d'autant que la sophistication et le volume des attaques ne cessent de croître. Dans ce contexte, les approches basées sur l'apprentissage profond (deep learning) deviennent de plus en plus attractives et retiennent l'attention des chercheurs et des experts du domaine. En effet, dans le chapitre 2, nous introduisons un état de l'art de ces approches.

Chapitre II :

Techniques de Deep Learning

Chapitre II : Techniques de Deep Learning

II.1. Introduction

Le Deep Learning (DL), ou apprentissage profond a connu des progrès importants réalisés ces dernières années, lui permettant d'apporter une réelle valeur ajoutée dans de nombreux domaines, y compris celui de la cybersécurité des infrastructures critiques. Son développement continue d'ailleurs d'avancer à grands pas, porté par l'augmentation des capacités de calcul, la disponibilité de grandes masses de données et l'amélioration continue des architectures neuronales. Les géants du secteur, comme Google, Meta, Microsoft ou NVIDIA, investissent massivement dans ce domaine, et bien de nombreuses applications sont désormais propulsées par le DL.

Dans ce chapitre, nous mettrons l'accent sur le concept de DL et les différentes techniques utilisées dans ce domaine. Nous présenterons successivement les architectures fondamentales (réseaux de neurones profonds, réseaux convolutionnels, réseaux récurrents), leurs principes de fonctionnement.

II.2. Deep Learning

II.2.1. Définition:

Le Deep Learning (apprentissage profond) est un sous-domaine du Machine Learning (ML), introduit afin de rapprocher ce dernier de son objectif principal : l'intelligence artificielle (figure II.1). Il regroupe des algorithmes inspirés de la structure et du fonctionnement du cerveau humain. Ces algorithmes sont capables d'apprendre plusieurs niveaux de représentation, ce qui leur permet de modéliser des relations complexes entre les données [35].

Le Deep Learning repose sur le principe des réseaux de neurones artificiels et est conçu pour traiter de grands volumes de données en ajoutant des couches successives au réseau. Un modèle d'apprentissage profond peut extraire automatiquement des caractéristiques à partir de données brutes grâce à ses multiples couches de traitement, composées de transformations linéaires et non linéaires. Il apprend progressivement ces caractéristiques à travers chaque couche, avec une intervention humaine minimale.

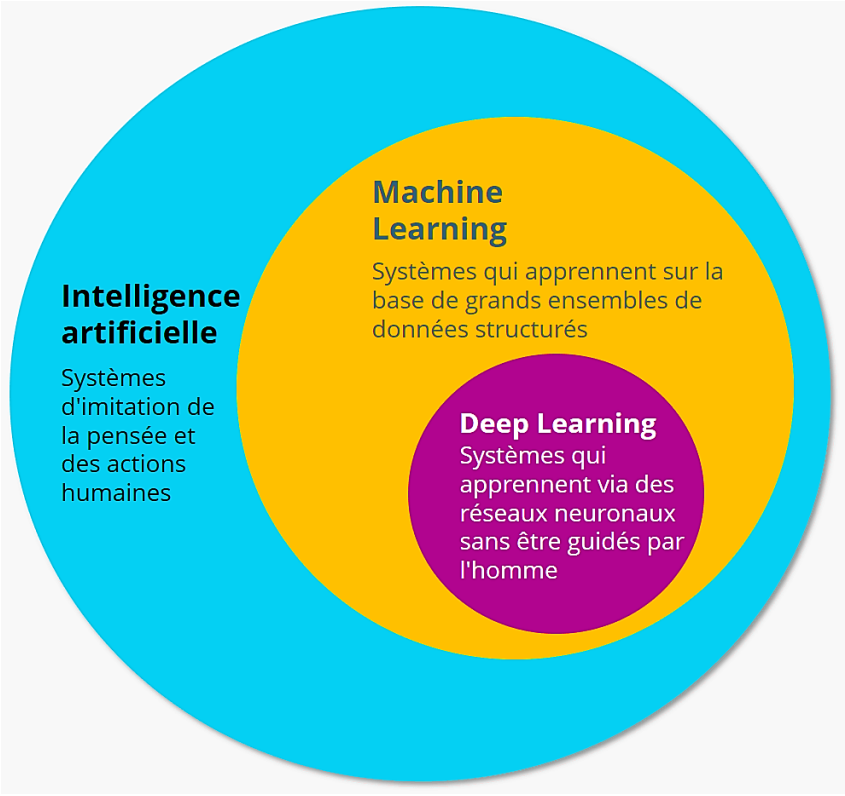


Figure II.1: La différence entre l'intelligence artificielle, machine learning et deep learning

II.2.2. Les avantages et les inconvénients de DL:

L'apprentissage profond (Deep Learning) est capable d'extraire, de traiter et d'apprendre des représentations de données dans leur forme originale, non modifiée et non étiquetée. Cette approche est très limitée, voire impossible, en apprentissage machine classique (ML), où il est nécessaire de sélectionner soigneusement les caractéristiques des données, ce qui requiert une grande compétence en ingénierie des représentations. L'avantage du Deep Learning sur le ML traditionnel réside dans sa capacité à comprendre les caractéristiques des données brutes sans extraction manuelle. Cela permet d'économiser du temps et des efforts, en particulier pour les très grands volumes de données. De plus, les résultats obtenus par l'apprentissage profond sont généralement meilleurs, plus précis et plus perfectionnés que ceux de l'apprentissage machine classique [36]. Les points forts peuvent être résumés en :

- Meilleurs résultats par rapport aux autres méthodes d'apprentissage machine
- Pas de développement manuel des caractéristiques ni d'étiquetage des données nécessaire
- Exécution efficace des tâches routinières, sans écarts de qualité
- Capacité à traiter des données non structurées

- Diversification des services visant à simplifier l'utilisation des réseaux de neurones artificiels

Néanmoins, l'apprentissage profond nécessite une énorme puissance de calcul, à la fois pour assurer la maintenance des réseaux de neurones artificiels et pour traiter la grande quantité de données requises. En effet, pour fonctionner correctement, le Deep Learning doit pouvoir s'appuyer sur un très grand nombre de données. Pamis les limitations du DL, on trouve [37] :

- Nécessite une grande puissance de calcul
- Développement des algorithmes d'apprentissage relativement long
- Nécessite une vaste base de données
- Davantage de données d'instruction initiales requises qu'avec d'autres méthodes
- Décisions difficilement, voire pas du tout, compréhensibles (effet boîte noire)

II.2.3. Principe de fonctionnement du deep learning :

Les neurones sont organisés en trois types de couches distinctes :

1. **Couche d'entrée (Input Layer) :** elle reçoit les données d'entrée X_i et les transmet à la première couche cachée.

2. Couches cachées (Hidden Layers) : elles effectuent des calculs mathématiques sur les entrées. L'un des défis majeurs dans la conception des réseaux de neurones est de déterminer le nombre de couches cachées, ainsi que le nombre de neurones par couche.

3. Couche de sortie (Output Layer) : elle renvoie les données de sortie Y_i

Chaque connexion entre deux neurones est associée à un poids W_i. Ce poids détermine l'importance de la valeur d'entrée. Initialement, les poids sont définis de manière aléatoire.

Chaque neurone possède également une fonction d'activation. L'un de leurs objectifs est de normaliser la sortie du neurone. Une fois qu'un ensemble de données d'entrée a traversé toutes les couches du réseau, celui-ci renvoie les données de sortie via la couche de sortie.

Chaque paire de couches voisines est connectée . Les neurones d'une même couche (généralement appelés nœuds) n'ont aucune connexion entre eux. La figure II.2 illustre une architecture standard d'un modèle de réseau de neurones profond.

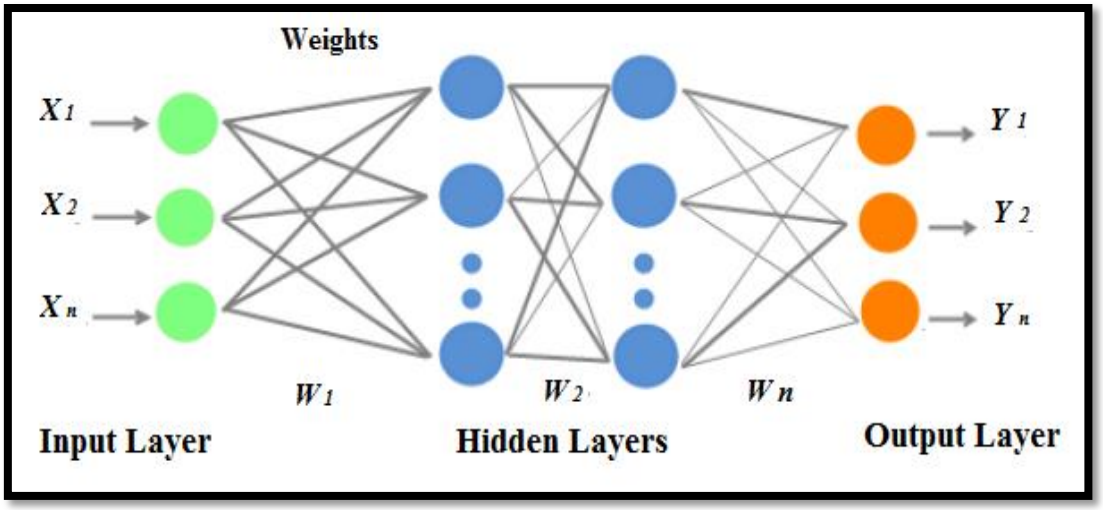


Figure II.2: Architecture d'un modèle Deep Learning

Fonction	Formule mathématique	Intervalle de sortie
Sigmoïde	$f(x) = \frac{1}{1+e^{-x}}$	[0, 1]
Tangente hyperbolique (Tanh)	$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$	[-1, 1]
ReLU (Rectified Linear Unit)	$f(x) = \max(0, x)$	[0, +∞[
Softmax	$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}$	[0, 1] (probabilités)

Tableau II.1: Principales fonctions d'activation

II.3. Les différentes architectures du deep learning:

Il existe différents algorithmes de Deep learning. Nous pouvons ainsi citer :

II.3.1. Convolutional Neural Networks (CNN) :

En apprentissage automatique, un réseau de neurones convolutifs ou réseau de neurones à convolution (en anglais CNN ou Convolutional Neural Networks) est un type de réseau de neurones artificiels acycliques (feed-forward), dans lequel le motif de connexion entre les neurones est inspiré par le cortex visuel des animaux. Les neurones de cette région du cerveau sont arrangés de sorte qu'ils correspondent à des régions qui se chevauchent lors du pavage du champ visuel. Leur fonctionnement est inspiré par les processus biologiques, ils consistent en un empilage multicouche de perceptrons, dont le but est de prétraiter de petites quantités

d'informations. Les CNN ont de larges applications dans la reconnaissance d'image et vidéo, les systèmes de recommandation et le traitement du langage naturel [38].

Les CNN sont composés de trois types de couches : des couches convolutives, des couches de regroupement et des couches entièrement connectées. Lorsque ces couches sont empilées, une architecture CNN a été formée. Une architecture CNN est illustrée à la figure II.3.

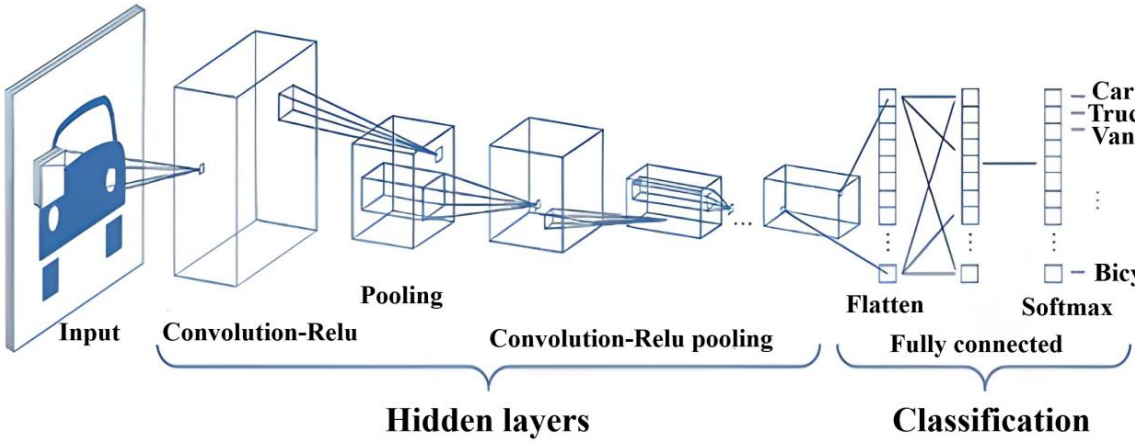


Figure II.3: Réseau de neurones CNN

- **Couche de convolution (CONV) :** La couche convolutive est le composant le plus important de toute architecture CNN. Il contient un ensemble de noyaux convolutifs (également appelés filtres), qui sont convolués avec la donnée d'entrée (métriques à N dimensions) pour générer une carte des caractéristiques.
- **Couche de pooling (POOL) :** Après chaque couche convolutive, il peut y avoir une couche de mise en commun. Cette couche est sous-échantillonnée le long de la dimension spatiale d'une entrée donnée, ce qui réduira encore le nombre de paramètres dans l'activation. Il existe plusieurs façons d'effectuer cette mise en commun, comme la prise de valeur moyenne ou maximale, ou la combinaison linéaire de neurones dans le bloc [38].
- **Couches de correction (RELU) :** Il est possible d'améliorer l'efficacité du traitement en intercalant une couche entre les couches de traitement qui effectuera une fonction d'activation sur les signaux de sortie. Cette fonction force les neurones à renvoyer des valeurs positives.

- **Couches entièrement connectée (FC : Fully connected) :** Après plusieurs couches de convolution et de max-pooling, le raisonnement de haut niveau dans le réseau neuronal se fait via des couches entièrement connectées. Les neurones d'un canapé (module) entièrement connecté sont connectés à toutes les sorties du canapé précédent. Leurs fonctions d'activation peuvent ainsi être calculées à l'aide d'une multiplication matricielle suivie D'un décalage de polarisation.

- **Couches de perte (LOSS) :** La couche de perte spécifie comment l'entraînement du réseau pénalise la différence entre le signal attendu et le signal réel. Normalement, elle est le dernier canapé du réseau. Il est possible d'utiliser une variété de fonctions de perte adaptées à des tâches spécifiques. SOFTMAX est une fonction qui calcule la distribution de probabilité sur les classes de sortie [38].

II.3.2. RNN (Recurrent Neural Network) [39] :

L'idée derrière les RNN est d'utiliser des informations séquentielles. Dans un réseau neuronal traditionnel, on suppose que toutes les entrées (et les sorties) sont indépendantes les unes des autres. Mais pour de nombreuses tâches, c'est une très mauvaise idée. Si on veut prédire le prochain mot dans une phrase, il faut connaître les mots qui sont venus avant. Les RNN sont appelés récurrents car ils exécutent la même tâche pour chaque élément d'une séquence, la sortie étant dépendante des calculs précédents.

Une autre façon de penser aux RNN est qu'ils ont une mémoire qui capture l'information sur ce qui a été calculé jusqu'à présent. En théorie, les RNN peuvent utiliser des informations dans des séquences arbitrairement longues, mais dans la pratique, on les limite à regarder seulement quelques étapes en arrière. Figure II.4 illustre un RNN typique:

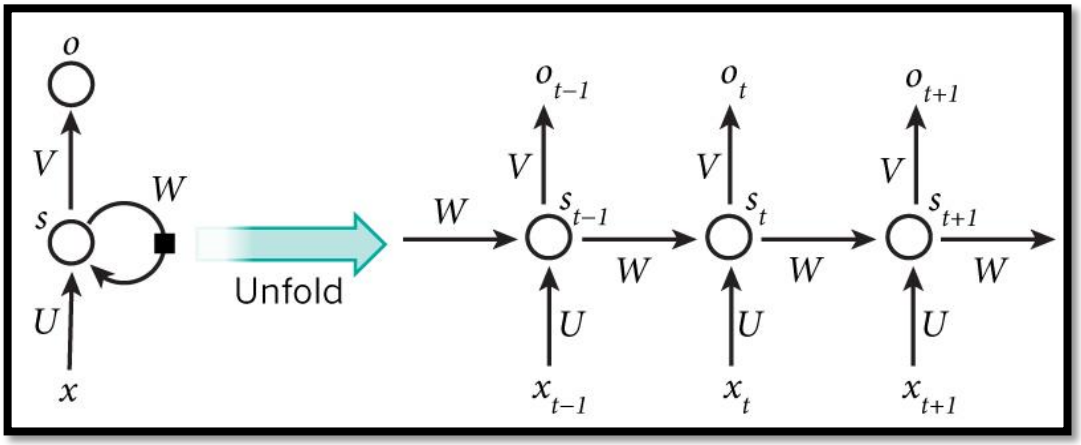


Figure II.4: Un RNN (à gauche). Sa version déroulée (à droite)

Le schéma ci-dessus montre un RNN déroulé. En déroulant, ça signifie simplement qu'on montre le réseau pour la séquence complète. Par exemple, si la séquence qui nous intéresse est une phrase de 5 mots, le réseau serait déroulé en un réseau de neurones de 5 couches, une couche pour chaque mot. Les formules qui régissent les calculs dans un RNN sont les suivantes :

- x_t est l'entrée au moment t.
- U, V et W sont les paramètres que le réseau va apprendre des données de l'apprentissage.
- s_t est l'état caché au moment t. C'est la mémoire du réseau. s_t est calculé en fonction de l'état caché précédent et de l'entrée à l'étape actuelle :

$$s_t = f(Ux_t + Ws_{t-1}) \quad \text{(II.1)}$$

- f est une fonction non linéaire telle que : Relu ou Hyperbolique tangent (tan).
- o_t est la sortie au moment$t$. Par exemple, si on veut prédire le prochain mot dans une phrase, ce serait un vecteur de probabilités dans un vocabulaire.

$$o_t = softmax(Vs_t) \quad \text{(II.2)}$$

On peut considérer l'état caché s_t comme la mémoire du réseau. s_t Capture des informations sur ce qui s'est passé dans toutes les étapes précédentes. La sortie o_t est calculée uniquement en fonction de la mémoire au moment$t$. Comme mentionné brièvement ci-dessus, c'est un peu plus compliqué dans la pratique car s_t ne peut généralement pas capturer des informations depuis trop longtemps.

Contrairement à un réseau de neurones traditionnel, qui utilise différents paramètres à chaque couche, un RNN partage les mêmes paramètres (parameters sharing) (ici :U, V, W) à travers toutes les étapes. Cela reflète le fait qu'on effectue la même tâche à chaque étape, juste avec des entrées différentes. Ce qui réduit le nombre total de paramètres que le réseau devra apprendre.

Le schéma ci-dessus présente des sorties à chaque moment, mais selon la tâche, ceci peut ne pas être nécessaire. Par exemple, lors de la prédiction du sentiment véhiculé par phrase, on ne peut. On soucier que de la sortie finale, pas du sentiment après chaque mot. De même, on ne peut pas avoir besoin d'entrées à chaque étape.

II.3.3. Les réseaux de mémoire à long terme à court terme (LSTM):

Un LSTM est un bloc fonctionnel spécial de réseaux neuronaux récurrents (RNN) doté d'une mémoire à court terme à long terme. Il s'agit d'une évolution des RNN et permet de résoudre le problème des gradients disparaissant, dans lequel les gradients de poids diminuent progressivement pendant l'entraînement et le réseau ne stocke donc plus d'informations utiles. Les cellules LSTM ont trois types de portes - une porte d'entrée, une porte de mémorisation et d'oubli et une porte de sortie.

Pour stocker les souvenirs des expériences passées. La mémoire à court terme est conservée longtemps et le comportement du réseau est encodé dans les poids. Les réseaux LSTM sont particulièrement adaptés pour faire des prédictions basées sur des données de séries temporelles, comme par exemple pour la reconnaissance de textes manuscrits et la reconnaissance vocale.[40]

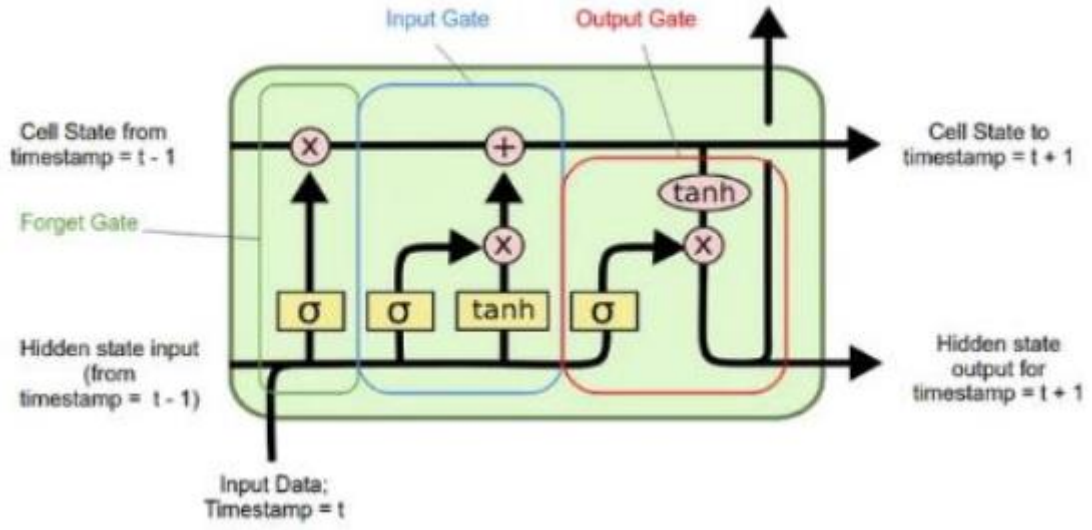


Figure II.5: Architecture de LSTM

Les portes d'un LSTM sont analogiques, sous la forme de sigmoïdes, ce qui signifie qu'elles vont de 0 à 1. Le fait qu'elles soient analogiques leur permet de faire une rétropropagation avec elles.

Les problèmes posés par la disparition des gradients sont résolus grâce à LSTM, car les gradients sont assez raides et l'entraînement est relativement court et la précision élevée.[40 36] La cellule mémoire d'un LSTM est composée de plusieurs portes : une porte d'entrée, une porte

de sortie et une porte d'oubli. Ces portes régulent le flux d'informations à l'intérieur de la cellule mémoire, permettant ainsi de contrôler les informations à retenir et celles à oublier.

Cela donne aux LSTM la capacité de mémoriser des informations importantes sur de longues séquences et d'ignorer les éléments moins pertinents. h_{-t} est l'état caché habituel des RNN mais dans les réseaux LSTM nous ajoutons un deuxième état appelé C_t. Ici, htreprésente la mémoire courte du neurone et Ctreprésente la mémoire à long terme [40].

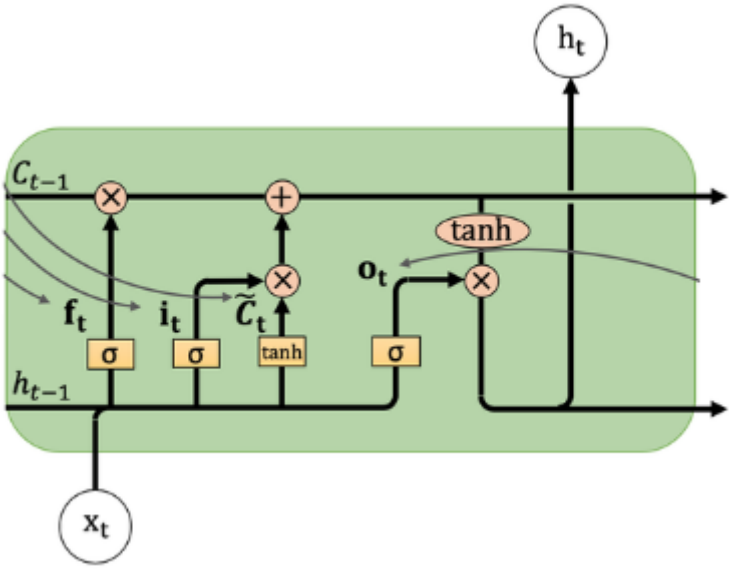


Figure II.6: Bloc de mémoire LSTM

La Porte entréeit (eq. II.3) détermine quelles information doivent être mises à jour dans la cellule mémoire. Elle prend en compte l'entrée actuel x_t ainsi que l'état précédent de la cellule mémoire h_{t-1} et génère un vecteur d'activation qui représente les informations à ajouter à la cellule $\hat{C}_{t-1}$ Cet ajout d'informations se traduit par une opération mathématique effectuée entre ce vecteur d'activation et l'état précédent $\hat{C}_{t-1}$ (eq. II.4)

$$i_t = \sigma(w_i . [_{ht-1}, x_t] + b_i) \quad \text{(II.3)}$$

$$\hat{C}_t = tanh(w_c, [h_{t-1}, x_t]) + b_i \quad \text{(II.4)}$$

La porte d'oubli f_t (eq.2.5) permet au LSTM de supprimer les informations obsolètes de la cellule mémoire. Elle utilise à la fois l'entrée actuelle et l'état précédent pour générer un vecteur d'activation qui détermine quelles informations doivent être oubliées. Cela se traduit également par une opération mathématique effectuée entre ce deuxième vecteur d'activation et l'état précédent $\hat{C}_{t-1}$.C'est à partir des deux opérations précédentes sur $\hat{C}_{t-1}$que l'on obtient l'état actuel c_t (eq. 2.6).

$$f_t = \sigma(w_f . [h_{t-1}, x_t] + b_f) \quad \text{(II.5)}$$

$c_t = f_t . C_{t-1} + i_t . \hat{C}_t$ (II.6)

Enfin, la porte de sortie (3) (II.7) détermine la sortie du LSTM à un instant donné. Elle utilise l'entrée actuellext et l'état actuel de la cellule mémoire pour générer c_t un vecteur d'activation qui représente la sortie du LSTM. C'est ainsi que l'on obtient h_t(2.8) .

$o_t = \sigma(w_o . [h_{t-1}, x_t] + b_o)$ (II.7)

$h_t = o_t . tanh(c_t)$ (II.8)

où .et + représentent la multiplication et l'addition point par point et b_i, b_f, et b_o représentent les Bais.

La combinaison de ces trois portes permet au réseau LSTM de gérer efficacement les dépendances à long terme. Lors de la rétropropagation du gradient, les LSTM peuvent maintenir un flux d'informations constant à travers le temps, évitant ainsi le problème du vanishing gradient et permettant un apprentissage plus stable et plus précis.

II.3.4. Gated Recurrent Unit (GRU):

L'unité récurrente fermée GRU (Gated Recurrent Unit) a été introduite en 2014 par Cho et Al [41] pour résoudre le problème de disparition du gradient rencontré par les réseaux récurrents classiques mais aussi pour proposer une architecture avec moins de paramètres à entrainer par rapport à une LSTM. À l'instar de LSTM, l'unité GRU est l'élément de base d'une architecture GRU. Une passe avant de l'unité GRU est modélisé par les équations :

$z_t = \sigma(W_z . x_t + U_Z . H_{t-1} + b_z)$ (II.9)

$r_t = \sigma(W_r . x_t + U_R . H_{t-1} + b_r)$ (II.10)

$H_t = \tanh(W_H . x_t + U_{H(r_t . H_{t-1})} + b_h)$ (II.11)

$H_t = (1 - z_t) . H_{t-1} + z_t . H_t$ (II.12)

où σ est la fonction sigmoïde, z_t est le vecteur d'activation de la porte de mise à jour, r_t le vecteur d'activation de la porte de réinitialisation, h_t est le vecteur candidat et h_t est le vecteur output de l'unité GRU. W et U sont des poids, b est le vecteur biais (poids et biais sont à entraîner durant le processus d'apprentissage) [41].

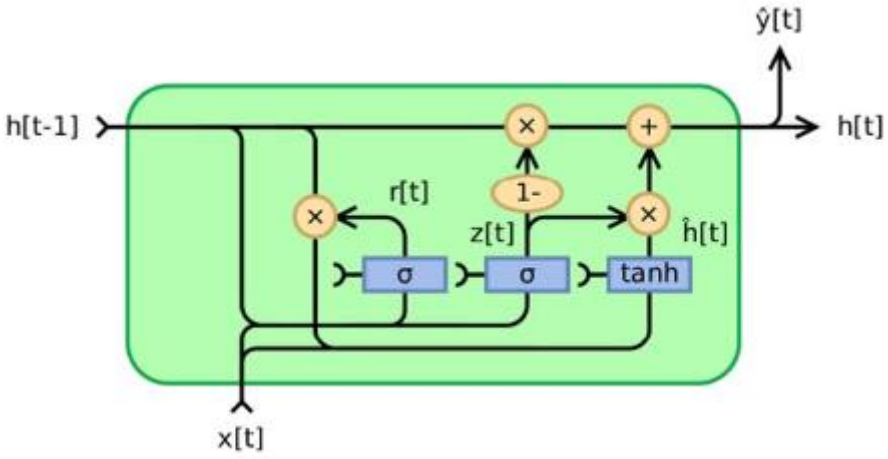


Figure II.7: Unité de base GRU.

II.4. DL pour la cybersécurité des infrastructures critiques

La disponibilité croissante de données issues des systèmes de contrôle industriels (SCADA), des réseaux de terrain, des automates programmables industriels (API/PLC) et des infrastructures de supervision, combinée à la nécessité de protéger des services essentiels (énergie, eau, transports, santé), a conduit à exploiter des méthodes avancées telles que l'apprentissage automatique, l'exploration de données, les statistiques et d'autres approches interdisciplinaires. Dans ce contexte, deep learning peut être appliqué aux systèmes de détection d'intrusions (IDS) spécifiques aux infrastructures critiques, qu'ils soient basés sur des signatures ou sur la détection d'anomalies comportementales.

Ces méthodes de classification et de prédiction permettent de détecter des schémas et des comportements inhabituels dans les flux de données industrielles, signalant ainsi des cyberattaques potentielles (FDIA, Ransomware, DOS…). Elles facilitent une cyber-réponse en temps réel, adaptée aux contraintes de continuité de service propres aux infrastructures critiques. Non seulement ces approches détectent les attaques en cours, mais elles contribuent également à anticiper d'éventuelles futures menaces.

Cependant, la collecte massive de données issues des capteurs, des actionneurs et des sources de données industriels soulève un problème de big data. Les opérateurs d'infrastructures critiques exigent des IDS offrant un taux de détection très élevé et un taux de fausses alarmes extrêmement faible, car une alarme erronée peut entraîner des arrêts de production coûteux ou des interruptions de service essentiel. Les approches par apprentissage profond, capables de traiter efficacement de très grands volumes de données tout en apprenant des modèles

complexes de normalité industrielle, sont donc particulièrement adaptées. Elles ont été introduites pour la détection d'anomalies au sein des réseaux industriels, dans le but de distinguer les comportements normaux (processus de production, cycles de maintenance) des comportements anormaux (attaques, dysfonctionnements suspects), permettant ainsi de détecter des activités malveillantes ou potentiellement dangereuses pour la continuité des services critiques.

II.5. Conclusion

Ce chapitre a posé les fondements théoriques nécessaires à la compréhension de l'apprentissage profond et de son applicabilité à la cybersécurité des infrastructures critiques. Après une introduction aux concepts essentiels du deep learning, nous avons détaillé les architectures neuronales les plus représentatives, nous avons décrit les CNN, RNN, LSTM et GRU

L'analyse des potentialités d'utilisation de ces modèles dans le contexte spécifique des infrastructures critiques a révélé plusieurs atouts majeurs : la capacité à traiter de grands volumes de données industrielles, la détection d'anomalies comportementales subtiles dans les processus physiques, et l'adaptation aux dynamiques temporelles complexes des systèmes IC et des réseaux de terrain. Toutefois, des défis subsistent, notamment la gestion des faux positifs, la contrainte du temps réel et l'interprétabilité des décisions.

Dans le chapitre suivant, nous exploitons ces bases théoriques pour explorer les performances de l'hybridation de ces modèles, combinant par exemple CNN pour l'extraction spatiale et LSTM ou GRU pour la modélisation temporelle, appliquée à la détection des cyberattaques dans les infrastructures critiques. L'objectif sera d'évaluer si ces architectures hybrides surpassent les modèles individuels en termes de précision, de taux de détection et de rapidité de réponse, face à des scénarios d'attaques réalistes.

Chapitre III :

Hybridation des Modèles d'Apprentissage Profond

Chapitre III : Hybridation des Modèles d'Apprentissage Profond

III.1. Introduction

Les systèmes de détection et de classification d'intrusions réseau basés sur le deep learning ont fait l'objet de nombreux travaux de recherche. En effet, les algorithmes DL ont récemment suscité un intérêt croissant de la part de la communauté de cybersécurité, car ils peuvent efficacement apprendre un modèle à partir de données complexes et non linéaires afin de classer des échantillons de trafic inconnus, qu'ils soient normaux ou malveillants. Cependant, les attaques ciblant les IC présentent des caractéristiques spécifiques qui rendent difficile leur détection par un seul type de modèle. Elles combinent souvent des anomalies spatiales et des anomalies temporelles. De plus, les exigences opérationnelles des IC (temps de réponse ultra-rapides, très faible taux de fausses alarmes, capacité à traiter de grands volumes de données issues de capteurs et d'actionneurs) nécessitent des approches plus puissantes et plus robustes que les modèles individuels.

C'est dans cette optique que nous présentons, dans ce chapitre, une exploration approfondie des performances de l'hybridation des modèles CNN, RNN, LSTM et GRU appliquée à la détection des cyberattaques dans les infrastructures critiques. L'hybridation vise à tirer parti des forces complémentaires de ces architectures. L'objectif principal de ce chapitre sera d'évaluer les performances architectures hybrides sous différents scénarios et conditions de cyberattaques (Reconaissance, DOS, MIMT et FDIA), en termes de plusieurs métriques critiques pour les infrastructures critiques : Accuracy, Recall, Precision, et F1-score. Nous confronterons ces modèles hybrides à des scénarios d'attaques réalistes, en nous appuyant sur des données représentatives des environnements IC.

III.2. Hybridation de modèle de Deep Learning

L'hybridation des réseaux de neurones vise à exploiter les forces complémentaires de différentes architectures. Dans cette section, nous présentons cinq modèles hybrides combinant des couches CNN et récurrentes RNN, LSTM, GRU pour la détection des cyberattaques dans les infrastructures critiques. Chaque modèle adopte une stratégie d'empilement spécifique selon que l'on privilégie l'extraction spatiale, la modélisation temporelle courte ou longue, ou encore la légèreté computationnelle.

III.2.1. CNN-LSTM

Ce modèle place les couches convolutionnelles en premier pour extraire les motifs spatiaux locaux des séquences de données réseau (paquets, trames, valeurs de capteurs). La LSTM prend ensuite modélise les dépendances temporelles à long terme entre ces motifs. Le traitement se déroule comme suit : la séquence d'entrée traverse deux blocs successifs composés chacun d'une couche Conv1D suivie d'un MaxPooling ; la sortie compressée est reformatée puis introduite dans une couche LSTM ; un Dropout régularise le modèle avant qu'une couche Dense ne projette le vecteur final vers la sortie (sigmoïde pour la classification binaire, softmax pour la classification multiclasse). La figure III.1 illustre cette architecture.

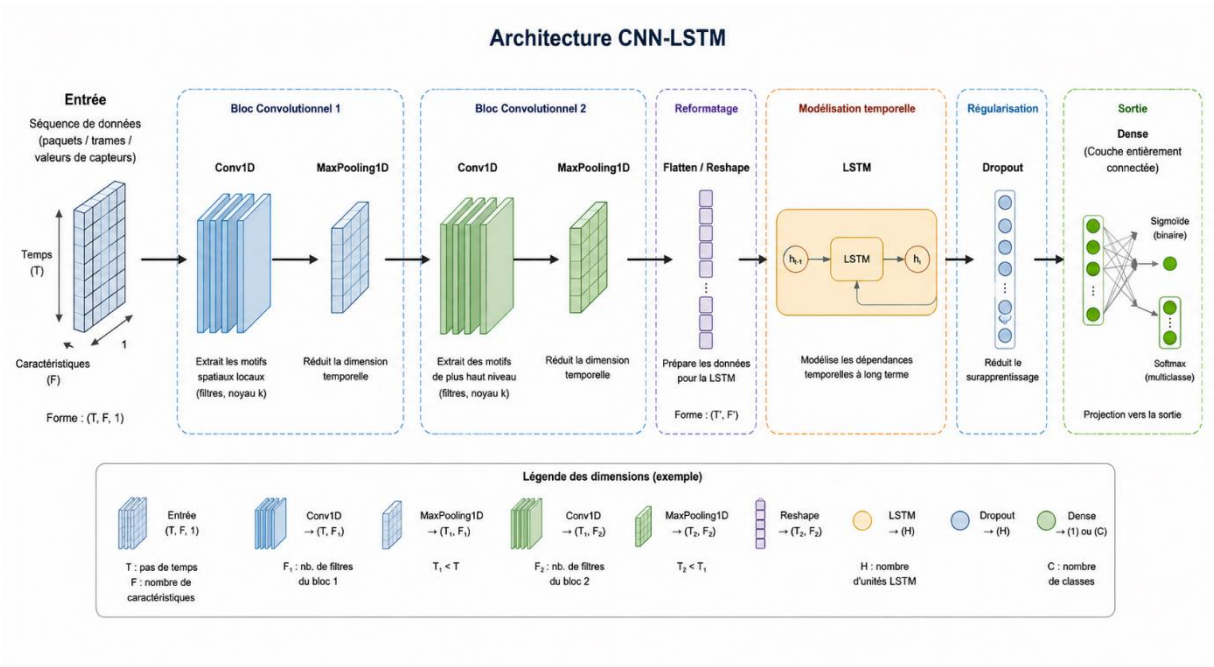


Figure III.1: Architecture du modèle CNN-LSTM

III.2.2. CNN-GRU

Ce modèle reprend le même principe d'extraction spatiale par couches CNN, mais remplace la LSTM par une GRU. Cette substitution rend l'architecture plus légère et plus rapide à entraîner, car la GRU comporte moins de paramètres (deux portes au lieu de trois, pas d'état de cellule distinct). Le déroulement est identique à celui du CNN-LSTM jusqu'à la sortie des couches convolutionnelles ; la séquence compressée est ensuite traitée par une couche GRU, suivie d'un Dropout, puis d'une couche Dense et de la couche de sortie. La figure III.2 présente cette architecture.

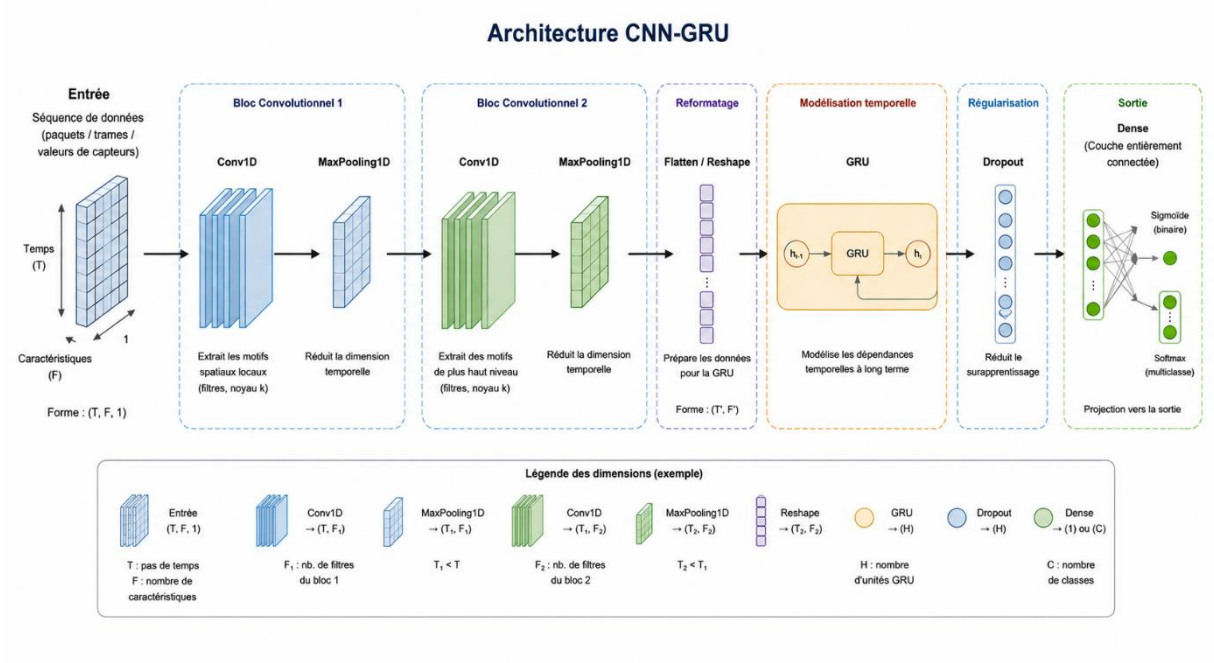


Figure III.2: Architecture du modèle CNN-GRU

III.2.3. RNN-LSTM

Contrairement aux deux modèles précédents, cette architecture n'utilise pas de couches convolutionnelles. Elle empile deux réseaux récurrents de complexité croissante. La RNN simple, par sa structure légère, capte d'abord les corrélations immédiates et les fluctuations brèves des données. La LSTM, qui lui succède, réinterprète cette représentation à une échelle temporelle plus large, grâce à sa mémoire cellulaire qui résiste à l'effacement progressif des gradients. Le traitement s'organise ainsi : la séquence brute alimente la couche RNN ; sa sortie est transmise à la couche LSTM ; le dernier état caché de la LSTM est projeté par une couche Dense vers la décision finale (sigmoïde ou softmax). La figure III.3 illustre cette architecture.

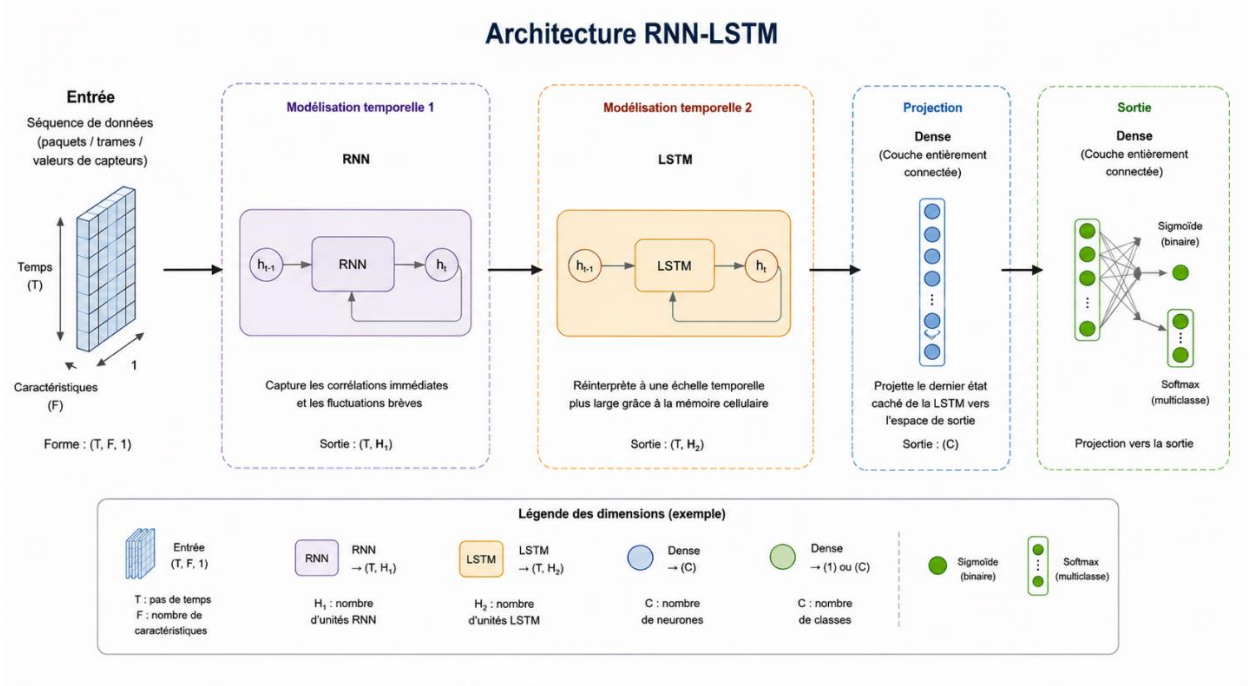


Figure III.3: Architecture du modèle RNN-LSTM

III.2.4. RNN-GRU

Ce modèle suit le même principe d'empilement que le RNN-LSTM, mais remplace la LSTM par une GRU. La RNN extrait d'abord une esquisse des variations instantanées, puis la GRU affine cette compréhension temporelle en appliquant ses deux portes (mise à jour et réinitialisation). La GRU offre un compromis intéressant : plus expressive qu'une simple RNN, mais plus économique qu'une LSTM. La séquence brute est d'abord présentée à la couche RNN ; la représentation obtenue est ensuite transmise à la couche GRU qui conserve sélectivement les informations passées pertinentes ; le vecteur final est projeté par une couche Dense vers la sortie sigmoïde ou softmax. La figure III.4 montre cette architecture.

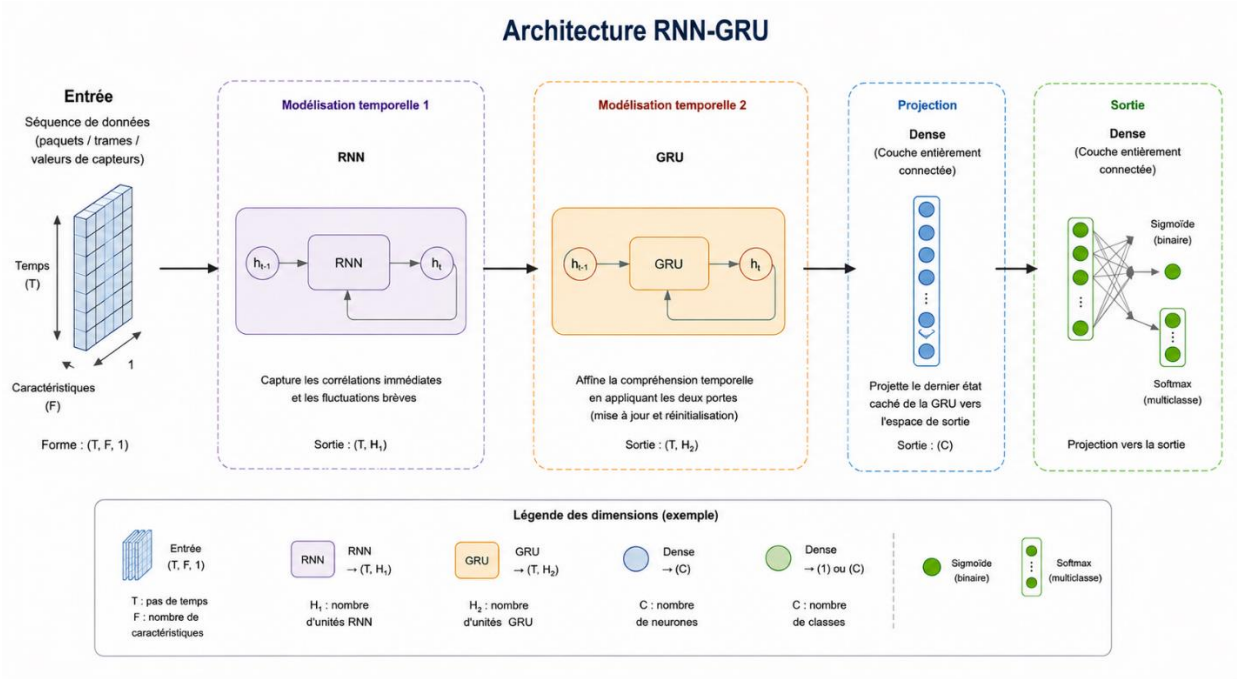


Figure III.4: Architecture du modèle RNN-GRU

III.2.5. LSTM-GRU

Cette architecture empile deux variantes de réseaux à portes, toutes deux capables de gérer les dépendances temporelles longues, mais avec des niveaux de complexité différents. La LSTM, placée en première position, exploite son état de cellule dédié et ses trois portes pour construire une représentation détaillée des dynamiques temporelles sur de longues périodes. La GRU, plus légère, prend ensuite le relais pour affiner cette représentation. Le traitement se déroule comme suit : la séquence brute est d'abord introduite dans la couche LSTM ; sa sortie est transmise à la couche GRU qui applique ses deux portes pour consolider l'information temporelle ; le vecteur final issu de la GRU est enfin projeté par une couche Dense vers la décision de classification (sigmoïde ou softmax). La figure III.5 illustre cette architecture.

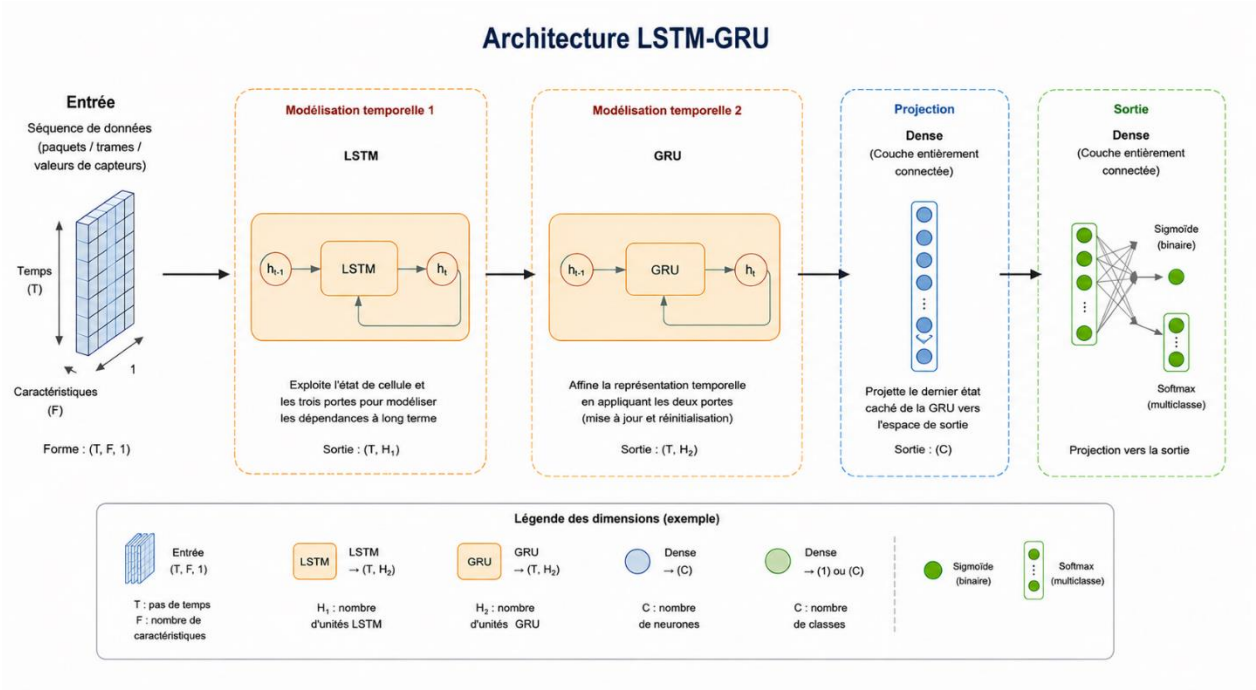


Figure III.5: Architecture du modèle LSTM-GRU

III.3. Procédure de classification des cyber-attaques par les modèles hybrides

La figure III.6 récapitule la méthodologie générale d'utilisation des modèles hybrides pour la classification des cyber-attaques dans les IC.

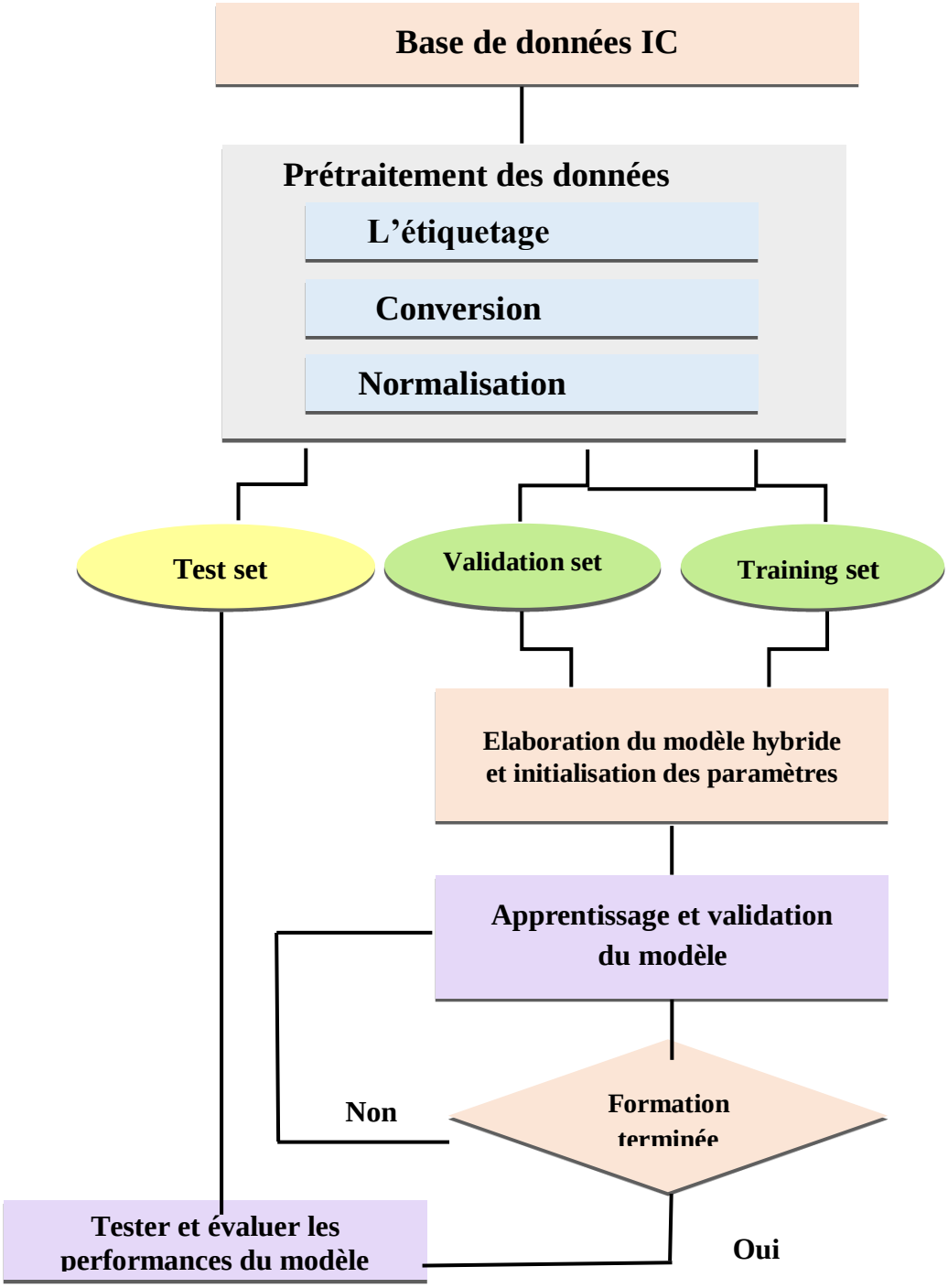


Figure III.6 : Schéma conceptuel de la classification des cyber-attaques par modèles hybrides

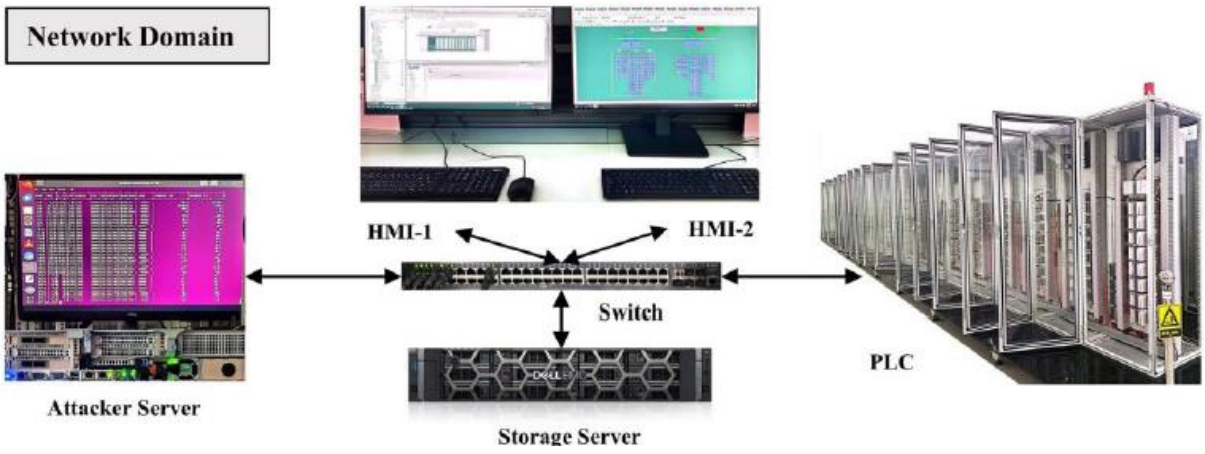
III.3.1 La base de données ICS-NAD [42,43]

La base de données ICS-NAD (Industrial Control Systems - Network Attack Detection) est un ensemble de données à grande échelle, collecté en environnement réel, conçu pour répondre aux limitations des précédents jeux de données sur la sécurité des ICS (tels que les petits bancs d'essai, les attaques non spécifiques aux ICS, et les caractéristiques insuffisantes). Il a été collecté sur un site d'essai industriel réel (figure III.6) contenant plusieurs marques d'ICS et processus industriels, incluant la production d'énergie thermique et le traitement des eaux usées.

Caractéristiques principales du dataset :

- **Environnement réel :** Les données ont été capturées à partir de trois marques d'ICS bien connues (ABB, Siemens, Schneider) utilisant trois protocoles réseau différents (TCP privé, S7Comm et Modbus).
- **Types d'attaques :** Il couvre 20 attaques réseau ICS courantes, regroupées en quatre catégories : Reconnaissance, DoS/DDoS, FDIA et MitM.
- **Modèles d'échantillonnage d'attaques :** Le dataset propose deux modèles distincts. Le premier présente un trafic d'attaque continu après une période normale. Le second, conçu pour les attaques DoS/DDoS, comporte des rafales d'attaque aléatoires de 30 secondes au sein de fenêtres d'une minute, répétées sur 20 cycles pour montrer les processus de déclenchement et de récupération du trafic.
- **Caractéristiques et étiquettes :** Il inclut 60 caractéristiques (basées à la fois sur les flux et sur les paquets) extraites à l'aide de l'outil ICSFlowGenerator. Tous les enregistrements sont clairement étiquetés comme "bénin" ou par le nom spécifique de l'attaque.
- **Taille et format :** La taille totale du dataset est de 245,96 Go, comprenant 272 fichiers de données. Il est fourni sous deux formats principaux : le trafic réseau brut (PCAP) et les caractéristiques extraites avec étiquettes (CSV).

a)



b)

Figure III.7: IC utilisé pour génération de la base de données ICS-NAD : a) Domaine physique (OT), b) Domaine réseau (IT) [43]

III.2. Environnement de programmation et ressources matérielles et logicielles :

III.2.1. Ressources matérielles :

Pour mener à bien notre travail, notre système est développé dans un micro-ordinateur ayant les caractéristiques techniques suivantes :

- Processeur : Intel ® Core™ i5-4590 CPU @3.30 Ghz 3.30 Ghz.
- RAM 8 Go.

III.2.2. Ressources logicielles :

- Système d'exploitation : Windows 10 64bit.
- Anaconda.

III.2.3. Environnement de programmation :

- Langage de programmation : python
- Editeur : Spyder 6.1.0

III.2.4. Bibliothèques utilisées :

Plusieurs bibliothèques ont été utilisées : Tensorflow, keras, numpy, pandas, matplotlib et sklearn.

Parmi toutes ces bibliothèques utilisées dans ce travail, la plus importante est bien Tensorflow. Cette dernière est une bibliothèque de logiciels open source destinée au calcul numérique de haute performance. Son architecture flexible permet un déploiement facile du calcul sur diverses plates-formes (processeurs CPU, processeurs graphiques GPU, processeurs tenseurs TPU), et des ordinateurs de bureau aux clusters de serveurs, aux périphériques mobiles et périphériques. La bibliothèque Tensorflow a été développée à l'origine par des chercheurs et des ingénieurs de l'équipe Google Brain au sein de l'organisation Google AI, elle prend en charge l'apprentissage automatique et l'apprentissage profond et le cœur de calcul numérique flexible est utilisé dans de nombreux autres domaines scientifiques.

III.3. Les mesures d'évaluation :

Pour comparer nos résultats, nous avons calculé les métriques lors de la phase de test, à savoir que la matrice de confusion, la précision, le recall et F1-score...etc. Ces métriques sont définies comme suit [44]:

- **Matrice de confusion :** est une matrice qui rassemble en lignes, les observations et en colonnes les prédictions. Les éléments de la matrice représentent le nombre d'exemples correspondants à chaque cas.

		Classe détectée (prédite)	
		Normal	**Attack**
Classe réelle	**Normal**	**TN (True Negative)**	**FP (False Positive)**
	Attack	**FN (False Negative)**	**TP (True Positive)**

Tableau III.1: Matrice de confusion

- → Un vrai négatif (TN) est une activité normale correctement classée lors de test.
- → Un faux positif (FP) est une activité normale mal classée pendant le test, c'est à dire, elle est considérée comme attaque.
- → Un faux négatif (FN) est une attaque non détectée pendant le test, c'est à dire, elle est considérée comme trafic normal.
- → Un vrai positif (TP) est une attaque correctement détectée.

- **Accuracy** ou le taux de réussite : c'est le rapport entre les enregistrements bien classés et la totalité d'enregistrements de test. Il sert à indiquer la façon dont la technique de détection est correcte.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{(III.1)}$$

- **Précision(Precision) :** appelée aussi le taux de reconnaissance, est la proportion de prédictions de positifs qui sont en effet des positifs.

$$Precision = \frac{TP}{TP+FP} \quad \text{(III.2)}$$

- **Recall :** c'est le taux des intrusions correctement détectées par rapport au nombre total d'intrusions. Il est calculé à partir de la formule suivante :

$$Recall = \frac{TP}{TP+FN} \quad \text{(III.3)}$$

- **F1-Score :** c'est une métrique qui combine la précision et le rappel en un nombre compris entre 0 et 1. Elle donne une évaluation de synthèse de la classification.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad \text{(III.4)}$$

III.4. Paramètres de simulations et configuration des modèles hybrides

Pour l'évaluation, chaque modèle est entraîné sur 60% de la base de données, validé sur 20%, et testé sur les 20% restants. La régularisation L2 ($\lambda = 10^{-4}$) et le dropout (0,5) sont appliqués systématiquement pour limiter le sur-apprentissage. Un early stopping avec une patience de 10 époques arrête l'apprentissage lorsque la perte de validation cesse de diminuer.

Les configurations des différents modèles, les paramètres de simulation sont récapitulé dans les tableaux III.2 et III.3.

Modèle	Couche CNN 1	Couche CNN 2	Couche Récurrente 1	Couche Récurrente 2	Couche Dense
CNN-LSTM	Conv1D(64,3,ReLU) + MaxPool(2)	Conv1D(128,3,ReLU) + MaxPool(2)	LSTM(100)	-	Dense(50, ReLU)
CNN-GRU	Conv1D(64,3,ReLU) + MaxPool(2)	Conv1D(128,3,ReLU) + MaxPool(2)	GRU(100)	-	Dense(50, ReLU)
RNN-LSTM	-	-	RNN(64, tanh)	LSTM(100)	Dense(50, ReLU)
RNN-GRU	-	-	RNN(64, tanh)	GRU(100)	Dense(50, ReLU)
LSTM-GRU	-	-	LSTM(100)	GRU(64)	Dense(50, ReLU)

Tableau III.2: Configuration des architectures hybrides

Paramètre	**Valeur**
Dropout	0,5
Sortie	Sigmoid (binaire)
Optimiseur	Adam
Taux d'apprentissage	10^{-3}
Fonction de perte	BCE (binaire)
Batch size	64
Epoch	256
Régularisation L2	10^{-4}
Early stopping	Patience = 10
Test et validation split	20%

Tableau III.3: Paramètres d'apprentissage et d'évaluation communs

III.5. Résultat et discussions

Nous avons évalué les cinq modèles hybrides (CNN-LSTM, CNN-GRU, RNN-LSTM, RNN-GRU, LSTM-GRU) sur le dataset ICS-NAD, en suivant les configurations présentées dans les tableaux III.2 et III.3. Chaque modèle a été entraîné pour classifier les flux réseau comme bénins ou comme appartenant à l'un des types d'attaques ICS. Nous présentons ci-dessous les résultats obtenus pour quatre attaques représentatives des différentes catégories : Reconnaissance, Dos/DDoS, FDIA et MIMT.

III.5.1. Attaque Portscan

Attaque de type reconnaissance, Portscan est un outil de scan UDP utilisé par l'attaquant pour détecter les ports ouverts sur les équipements cibles, permettant ainsi de cartographier le réseau industriel et d'identifier les vulnérabilités potentielles avant une attaque plus destructive. Les résultats obtenus sont reportés dans les figures III.7, III.8, .9 , III.10 et III.11.

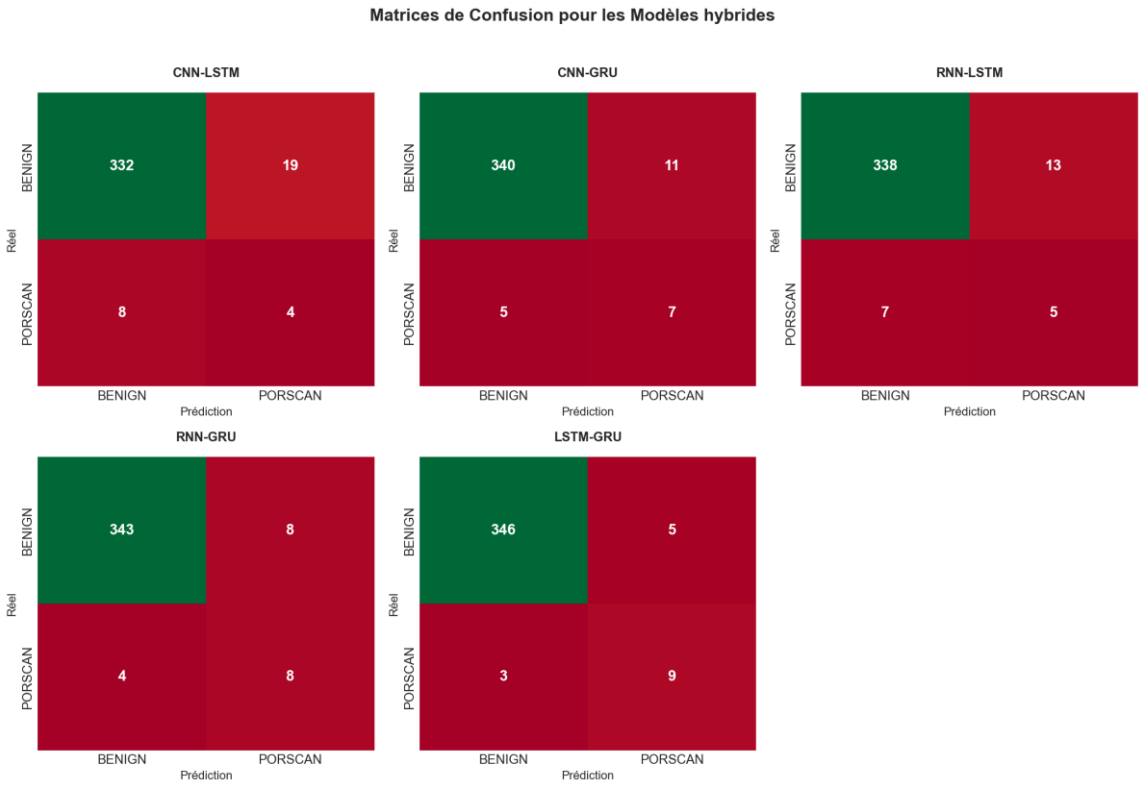


Figure III.8: Matrices de confusion des différents modèles hybrides

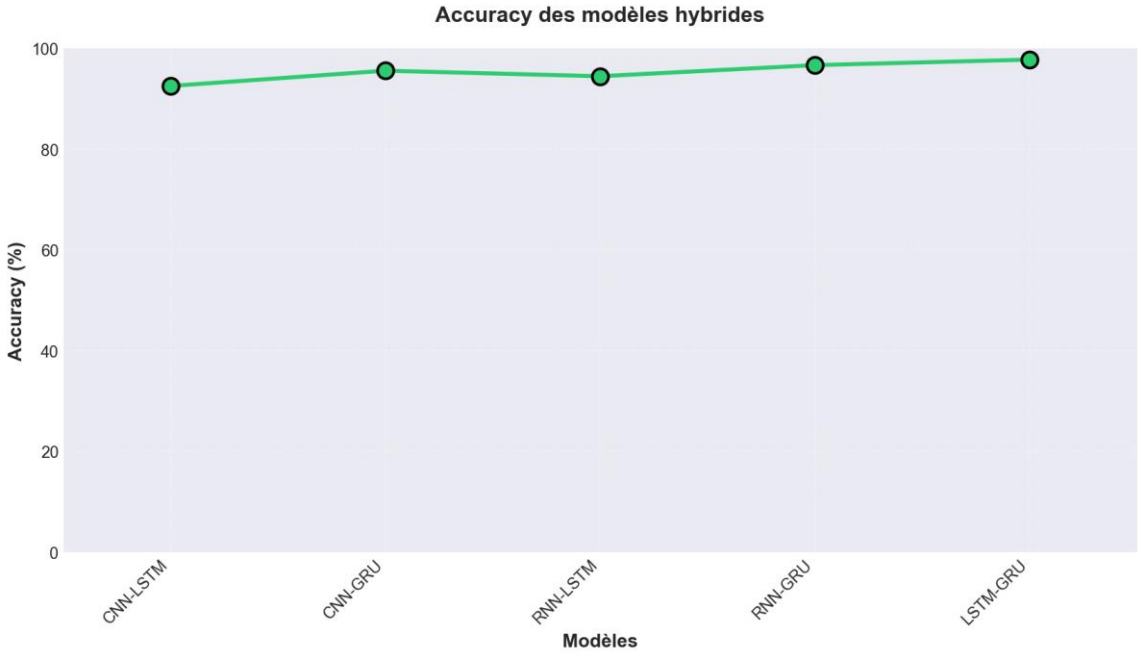


Figure III.9: Accuracy des différents modèles hybrides (attaque PORTSCAN)

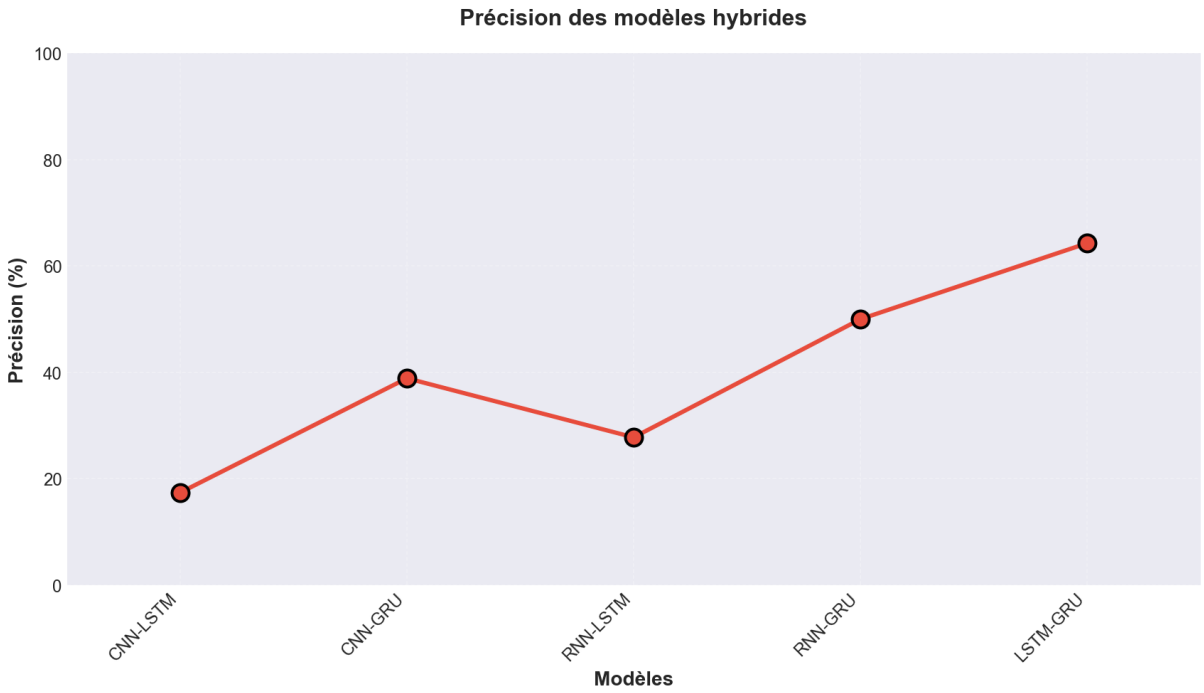


Figure III.10: Précision des différents modèles hybrides (attaque PORTSCAN)

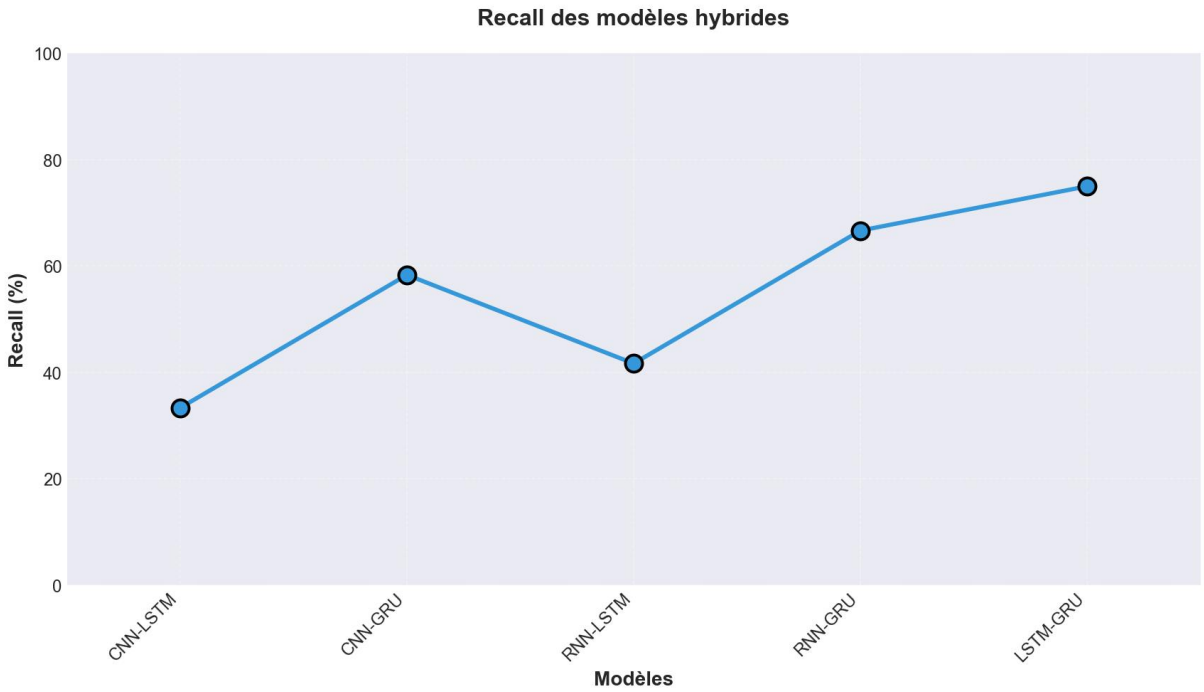


Figure III.11: Recall des différents modèles hybrides (attaque PORTSCAN)

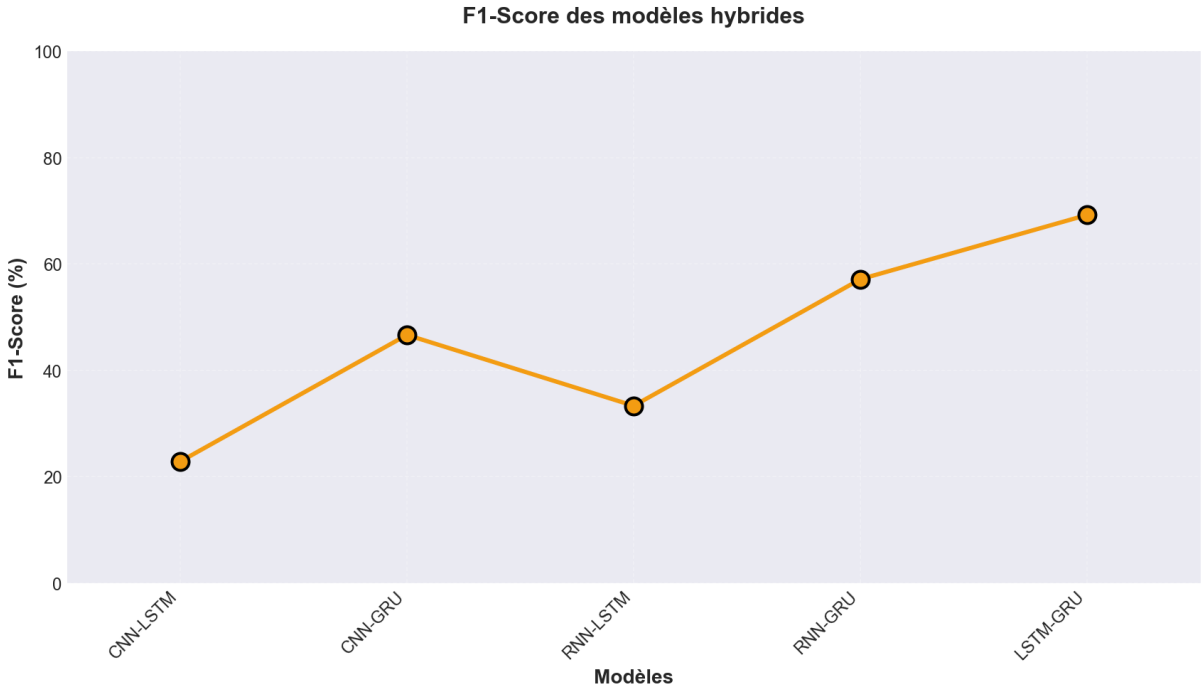


Figure III.12: F1-score des différents modèles hybrides (attaque PORTSCAN)

L'analyse des performances des différents modèles pour la détection des attaques PORSCAN révèle que LSTM-GRU offre les meilleures performances sur l'ensemble des métriques, avec un Recall de 75% (figure III. 10) et une précision de 64,29% (figure III.9), ce qui représente le meilleur compromis entre détection et fausses alertes. La détection des attaques PORSCAN (figure III.10) varie considérablement selon les modèles, allant de 33,33% pour CNN-LSTM à 75% pour LSTM-GRU, démontrant la supériorité des architectures hybrides LSTM-GRU pour ce type de détection temporelle. Les fausses alertes (BENIGN classés comme PORSCAN) oscillent entre 8 et 19 selon les modèles (figure III.7), ce qui représente un taux de faux positifs allant de 2,28% à 5,41% des BENIGN, avec LSTM-GRU produisant le moins de fausses alertes (8 FP). La précision pour la classe PORSCAN varie de 17,39% à 64,29%, cette faible précision s'explique principalement par le fort déséquilibre des classes (seulement 3,31% d'attaques), ce qui pénalise mécaniquement cette métrique. Le déséquilibre extrême des classes (96,69% BENIGN contre 3,31% PORSCAN) explique les performances plus faibles sur la classe minoritaire par rapport aux autres types d'attaques, mais LSTM-GRU parvient néanmoins à atténuer cet effet grâce à son architecture adaptée aux dépendances temporelles.

III.5.2. Attaque Teardrop

Cette attaque de type déni de service (DoS) exploite le mécanisme de réassemblage des paquets IP fragmentés en envoyant des fragments chevauchants aux offsets incorrects, ce qui perturbe le système cible et peut provoquer un plantage des équipements industriels comme les PLC ou les IHM. Nous avons obtenus les résultats illustrés sur les figures III.12, III.13, III.14 , III.15 et III.16.

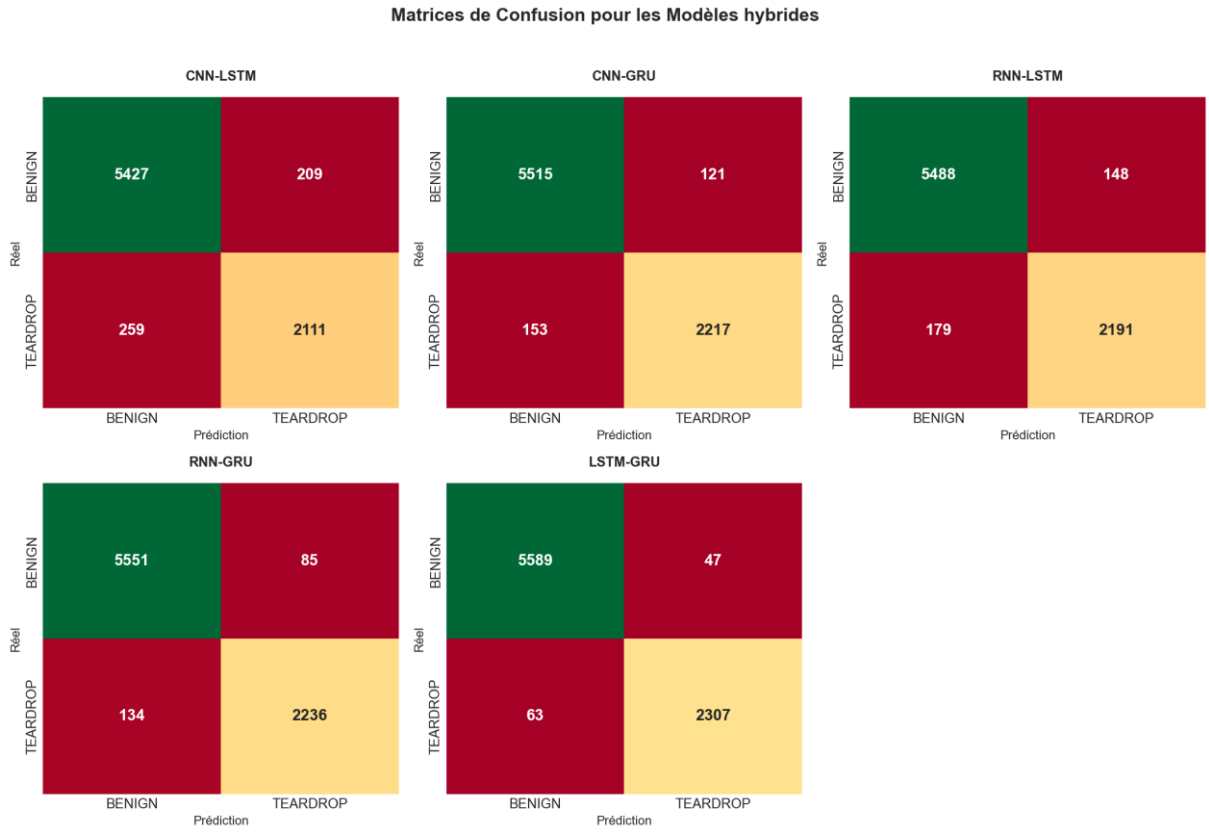


Figure III.13: Matrices de confusion des différents modèles hybrides (attaque TEARDROP)

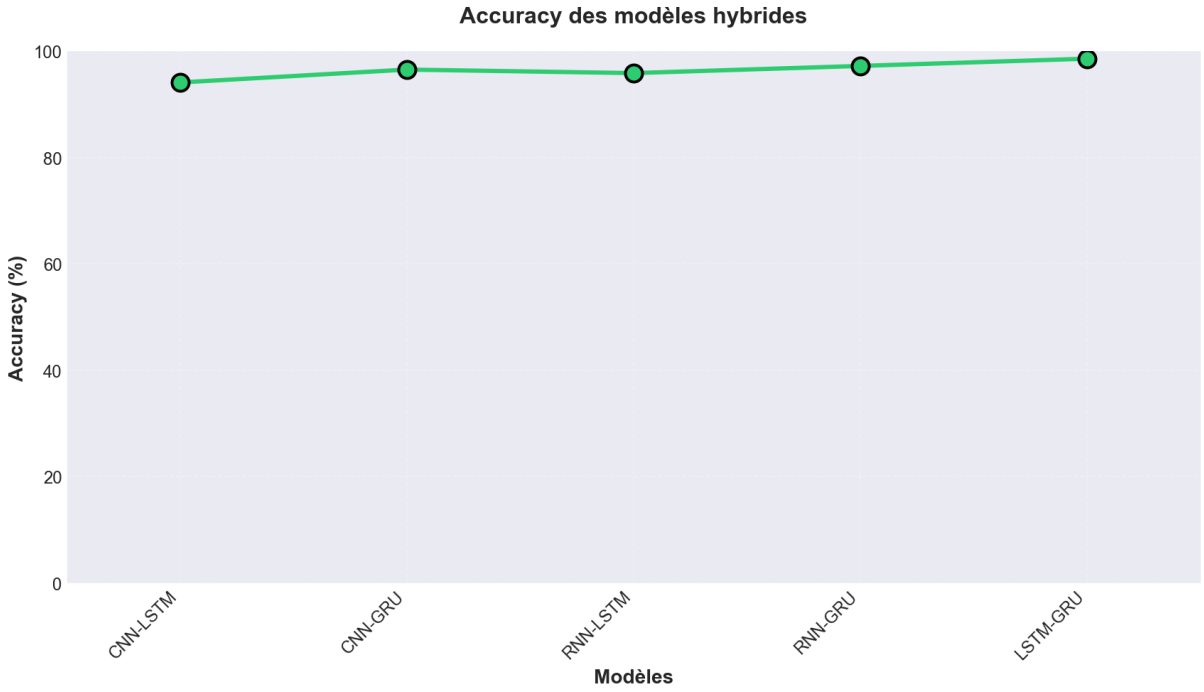


Figure III.14: Accuracy des différents modèles hybrides (attaque TEARDROP)

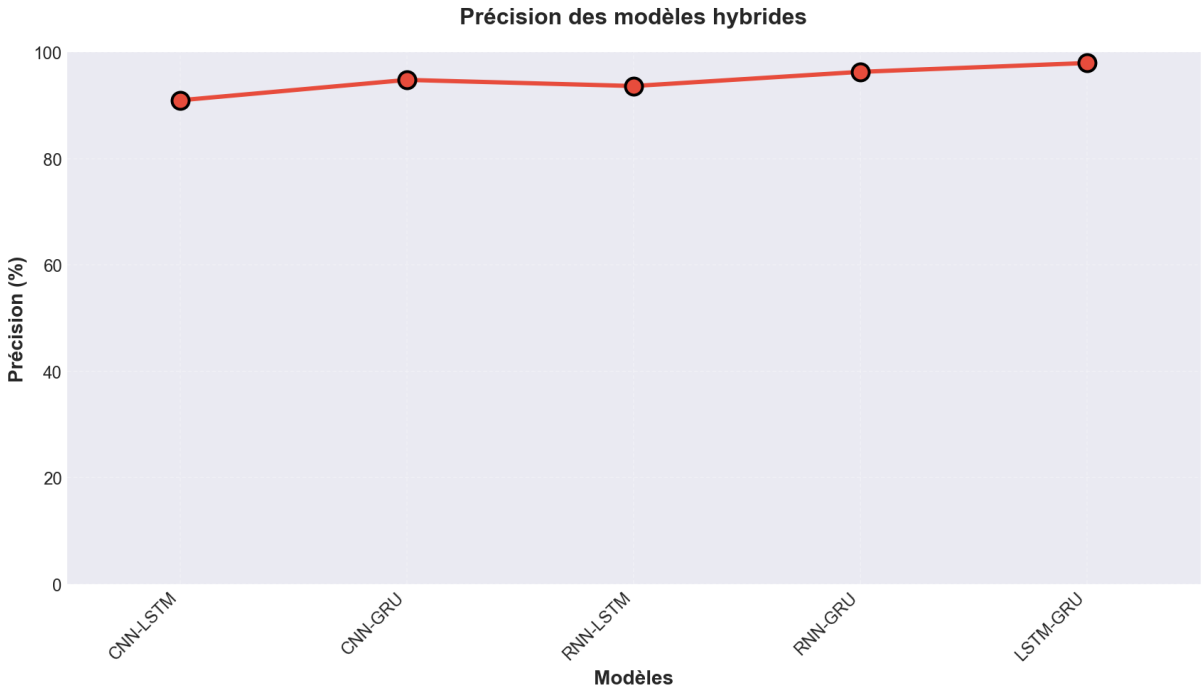


Figure III.15: Précision des différents modèles hybrides (attaque TEARDROP)

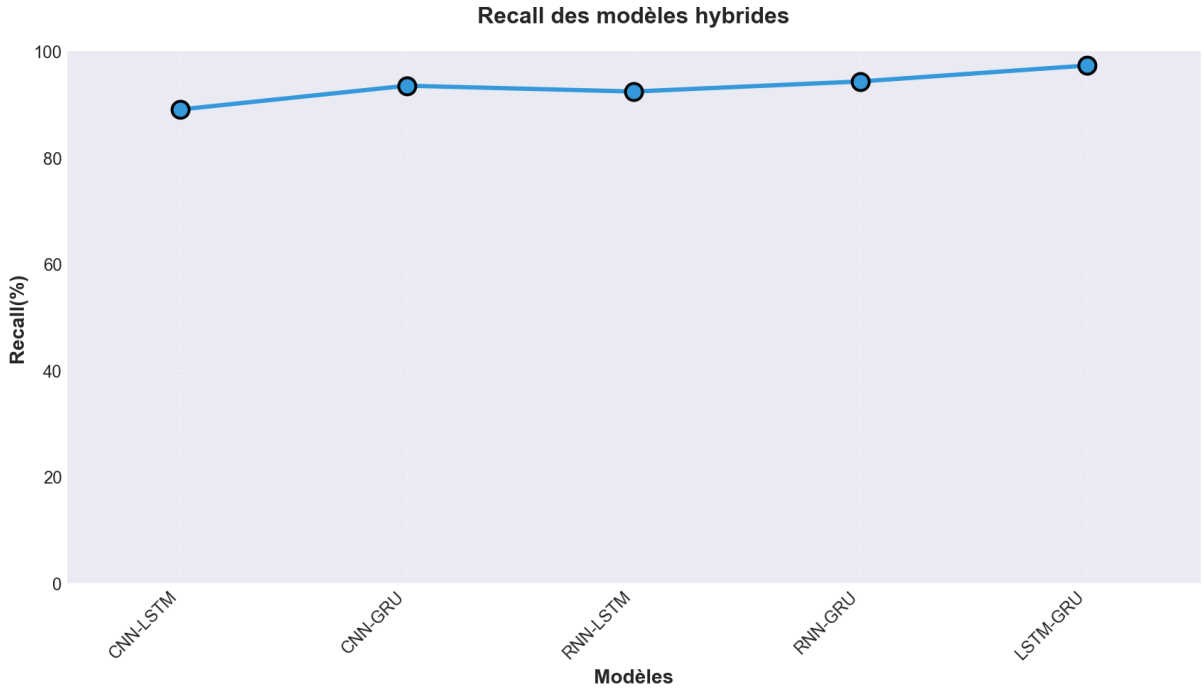


Figure III.16: Recall des différents modèles hybrides (attaque TEARDROP)

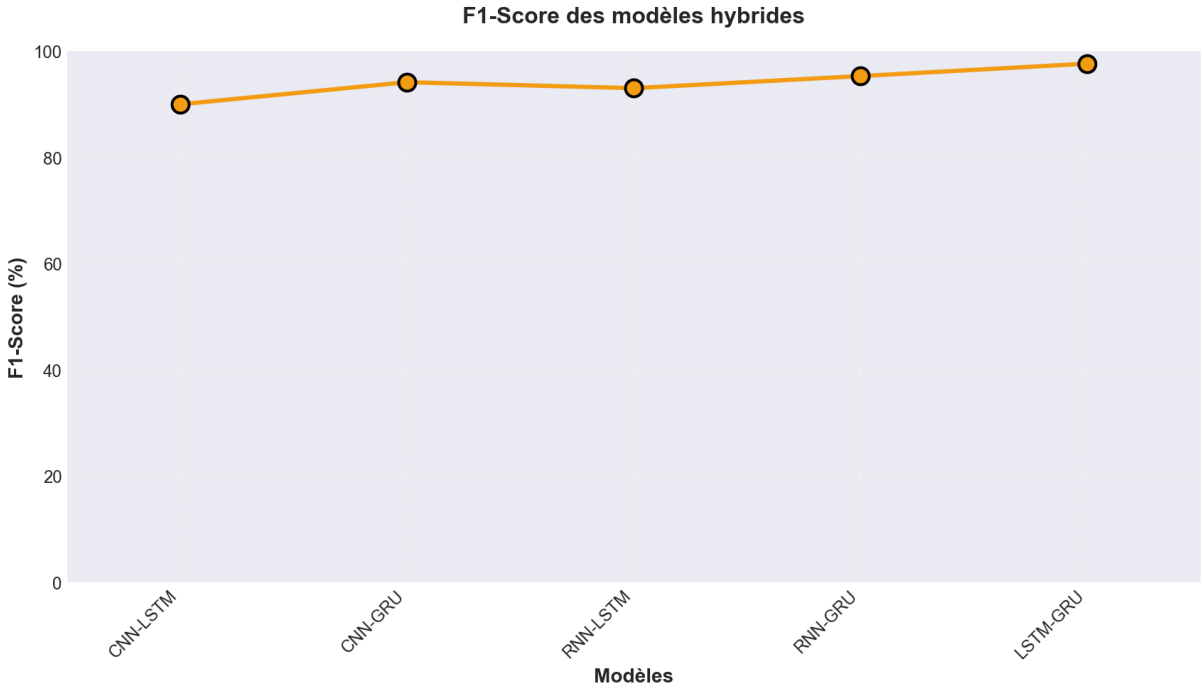


Figure III.17: F1-score des différents modèles hybrides (attaque TEARDROP)

L'évaluation des modèles pour la détection des attaques TEARDROP montre que LSTM-GRU domine toutes les métriques avec un recall de 97,34% (figure III.15) et une précision de 98,00% (figure III.14), soit une détection quasi-parfaite des 2307 attaques (figure III.12) . La détection des attaques TEARDROP est remarquable pour tous les modèles, avec des rappels allant de 89,07% (CNN-LSTM) à 97,34% (LSTM-GRU), ces performances élevées s'expliquent par un déséquilibre moins prononcé (29,60% d'attaques) comparé au dataset PORSCAN. Les fausses alertes varient de 47 à 209 selon les modèles, représentant un taux de faux positifs compris entre 0,83% et 3,71% des BENIGN, avec LSTM-GRU produisant seulement 47 FP (figure III.12). La précision (figure III.14) pour TEARDROP oscille entre 90,99% et 98,00%, tous les modèles atteignant une précision supérieure à 90%, ce qui confirme leur fiabilité pour cette classe d'attaque. Le déséquilibre modéré (70,40% BENIGN contre 29,60% TEARDROP) permet à tous les modèles d'atteindre des performances robustes, contrairement au cas PORSCAN où le déséquilibre extrême pénalisait la détection. Les architectures hybrides LSTM-GRU et RNN-GRU surpassent nettement CNN-LSTM, confirmant la supériorité des réseaux récurrents pour la détection d'attaques réseau temporelles.

III.5.3. Attaque ICMP REDIRECTION

Appartient aux attaques de type FDIA, l'ICMP REDIRECTION exploite un mécanisme légitime du protocole ICMP pour modifier frauduleusement la table de routage de la victime, détournant ainsi le trafic vers un hôte malveillant qui peut intercepter, modifier ou bloquer les communications entre équipements industriels. Cette attaque est plus subtile à détecter car elle utilise des paquets ICMP structurellement valides mais malveillants dans leur contenu. Les résultats obtenus sont présentés ci-dessous.

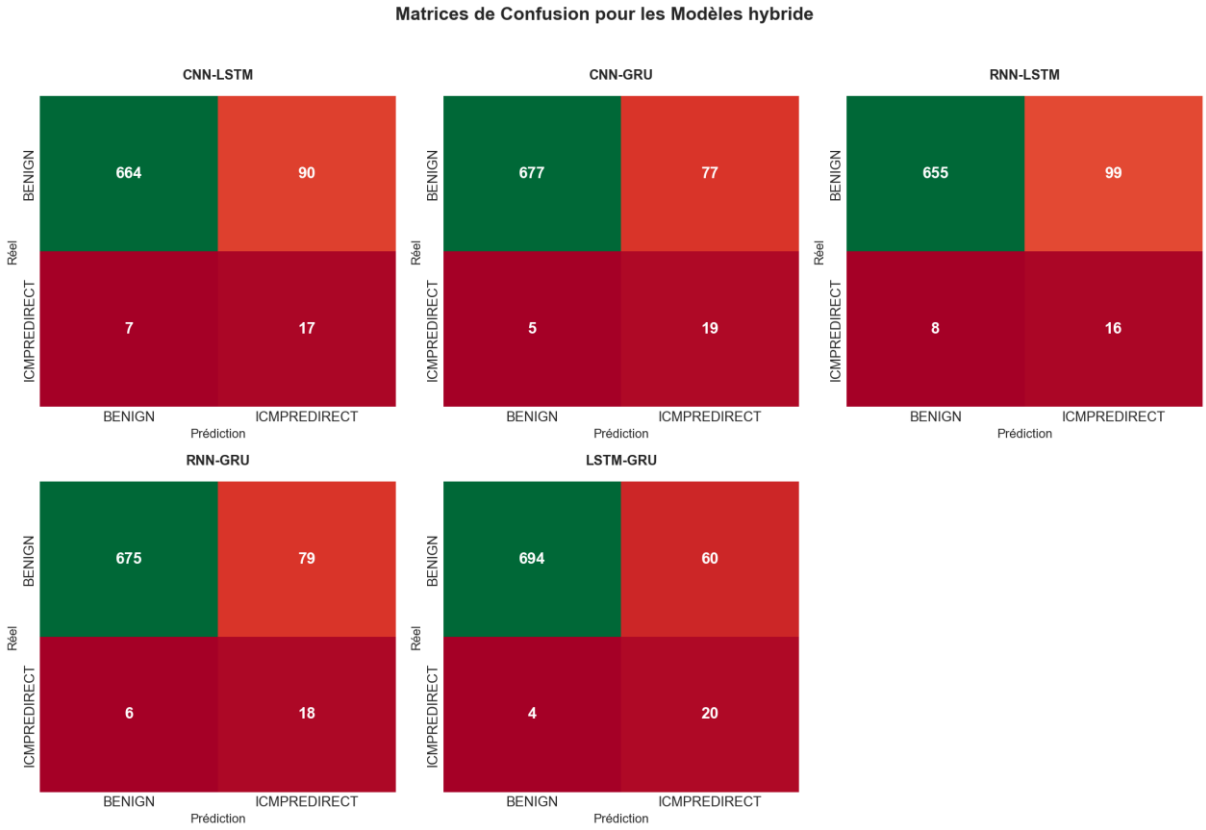


Figure III.18: Matrices de confusion des différents modèles hybrides (attaque ICMPREDIRECTION)

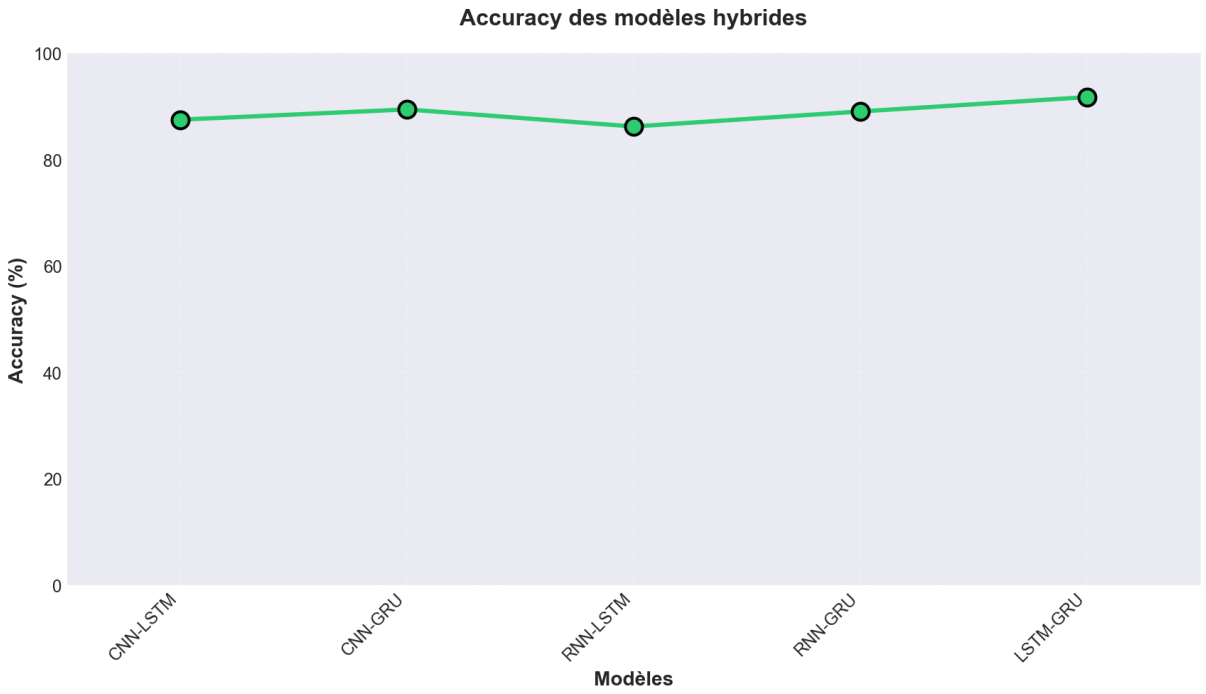


Figure III.19: Accuracy des différents modèles hybrides (attaque ICMPREDIRECTION)

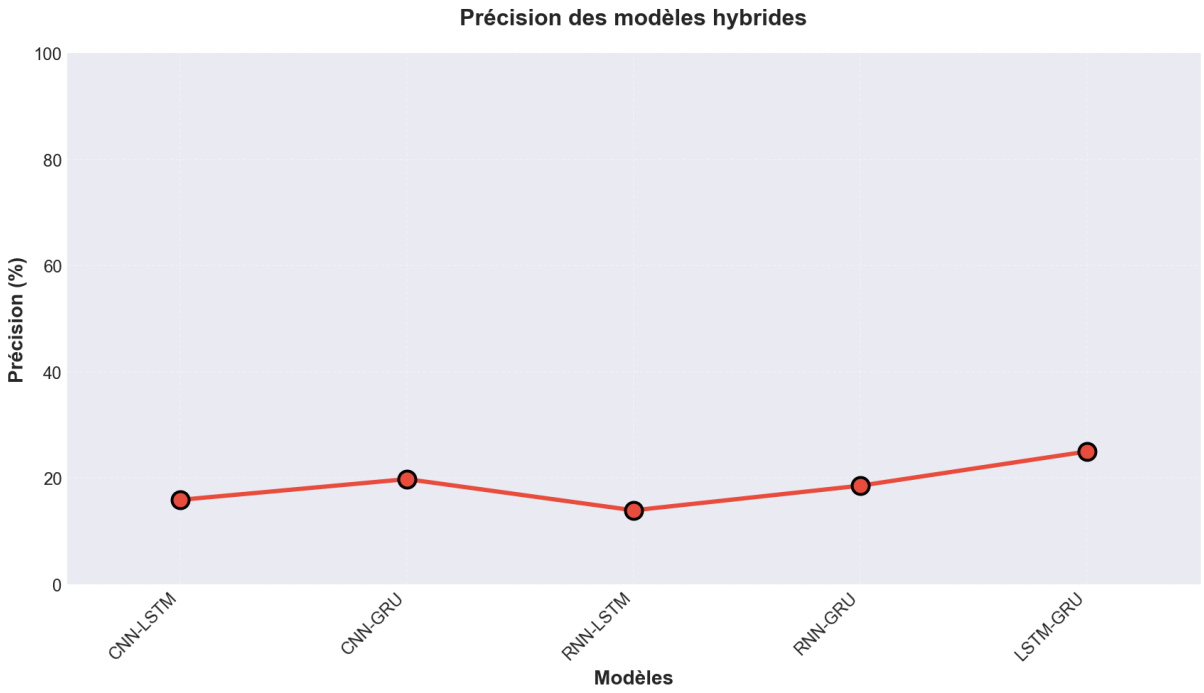


Figure III.20: Précision des différents modèles hybrides (attaque ICMPREDIRECTION)

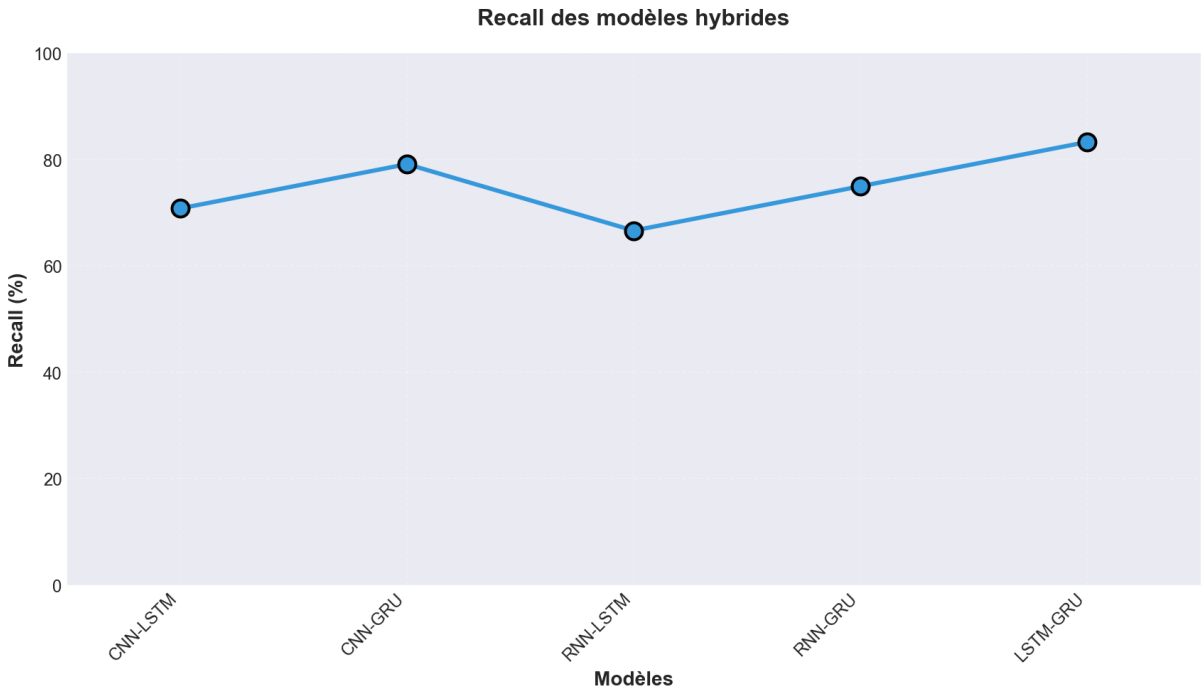


Figure III.21: Recall des différents modèles hybrides (attaque ICMPREDIRECTION)

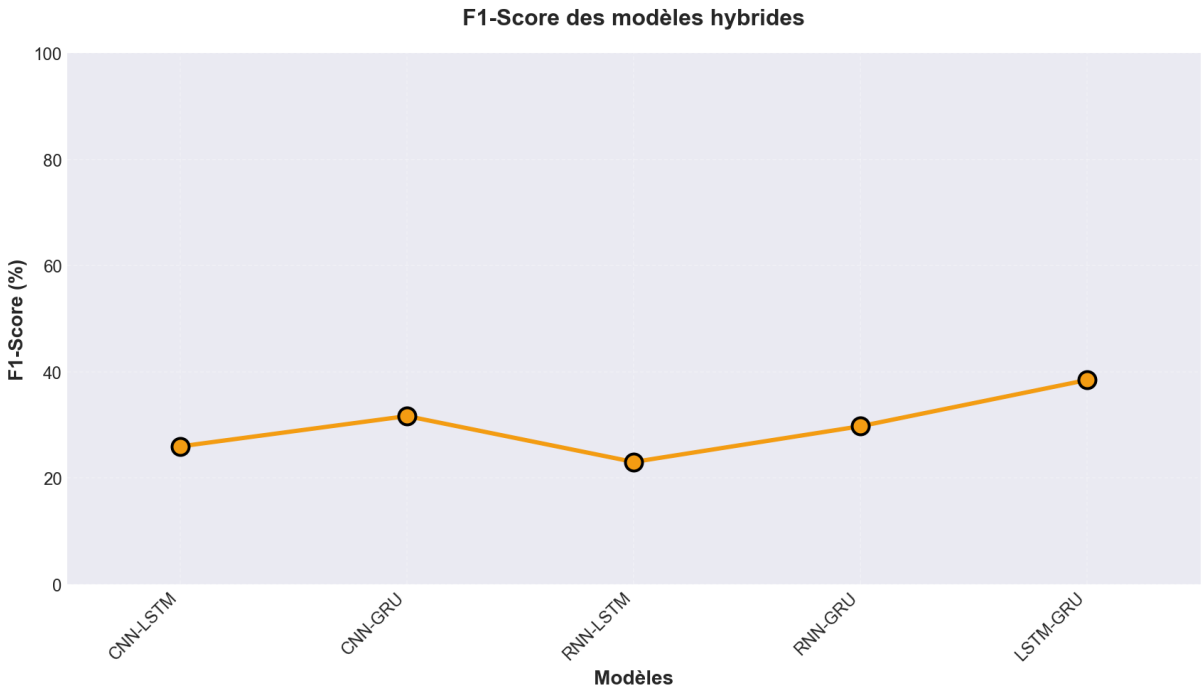


Figure III.22: F1-score des différents modèles hybrides (ICMPREDIRECTION)

L'analyse des performances pour la détection des attaques ICMPREDIRECT révèle que LSTM-GRU se distingue nettement avec un Recall de 83,33% (figure III.20) et une précision de 25,00%, (figure III.21) malgré un déséquilibre des classes extrêmement prononcé (seulement 3,08% d'attaques). La détection des attaques ICMPREDIRECT varie considérablement selon les modèles, allant de 66,67% (RNN-LSTM) à 83,33% (LSTM-GRU), ces performances modestes s'expliquent par la très faible représentation des attaques (24 échantillons sur 778). Les fausses alertes oscillent entre 60 et 99 (figure III.17) selon les modèles, représentant un taux de faux positifs allant de 7,96% à 13,13% des BENIGN, ce qui est relativement élevé et impacte la précision. La précision (figure III.19) pour ICMPREDIRECT varie de 13,91% à 25,00%, ces valeurs très faibles sont directement liées au fort déséquilibre des classes (96,92% BENIGN), car même un petit nombre de faux positifs dégrade considérablement cette métrique. Le déséquilibre extrême (96,92% BENIGN contre 3,08% ICMPREDIRECT) explique les performances limitées sur cette attaque, similaires au cas PORSCAN, contrairement au cas TEARDROP où le déséquilibre modéré permettait des détections bien supérieures. LSTM-GRU reste le modèle le plus performant, confirmant sa robustesse même dans des conditions de déséquilibre sévère.

III.5.4. Attaque ARP POISONING

Attaque de type MitM, l'ARP POISONING consiste à envoyer des réponses ARP falsifiées associant l'adresse MAC de l'attaquant à l'adresse IP d'une autre machine, ce qui place l'attaquant en position d'intercepter, consulter et modifier tout le trafic entre deux équipements industriels, par exemple entre un capteur et son PLC. Les résultats obtenus sont reportés dans la figure ci-dessous.

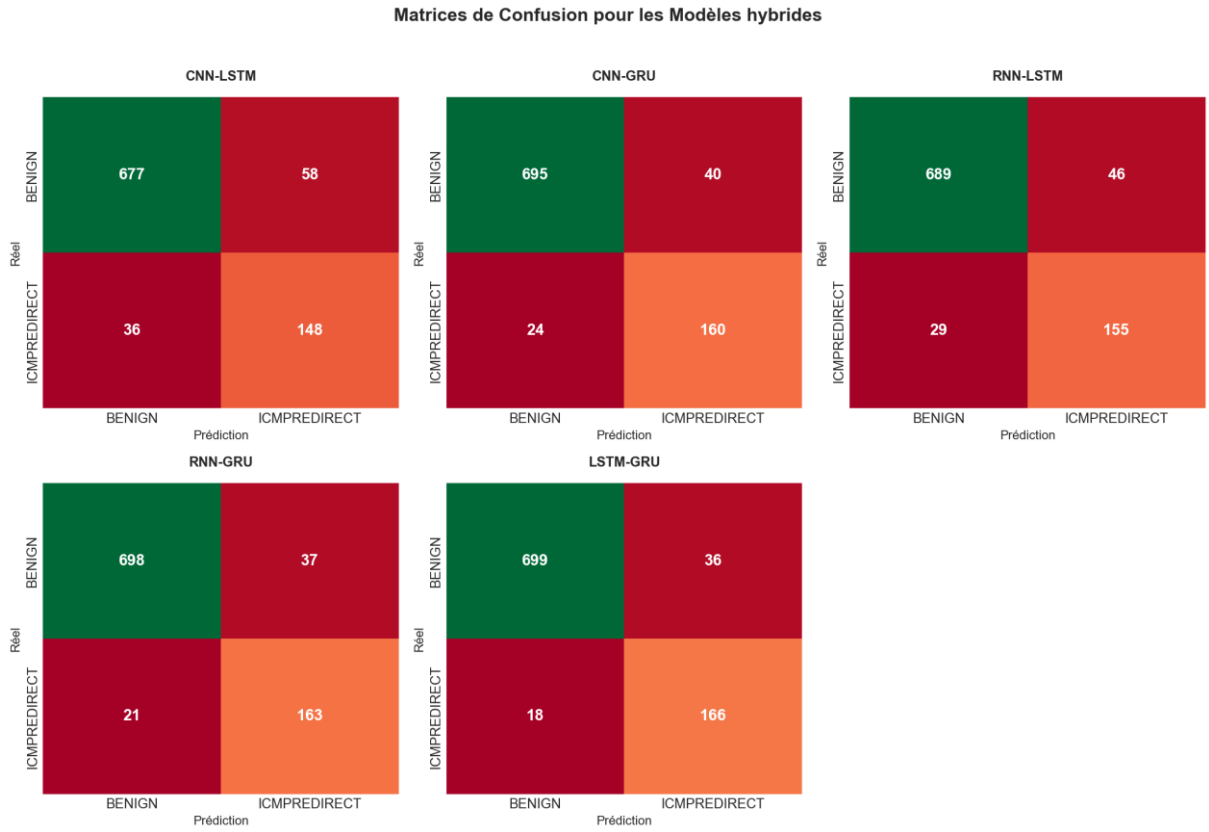


Figure III.23: Matrices de confusion des différents modèles hybrides (attaque ARP POISONING)

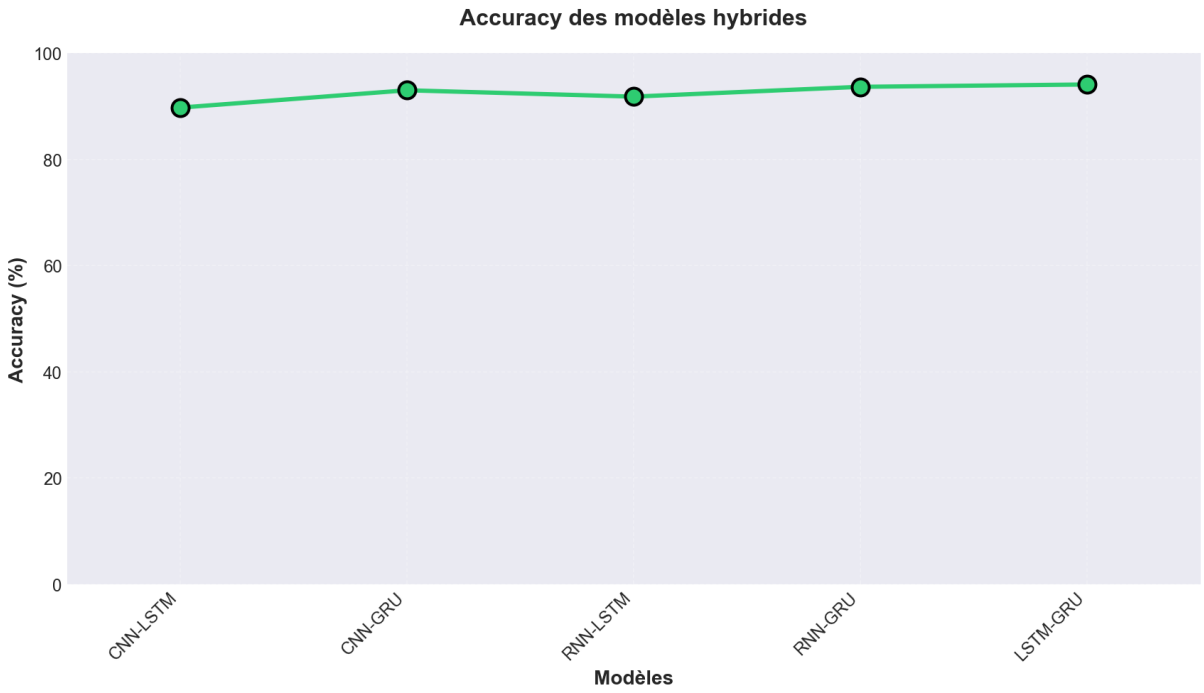


Figure III.24: Accuracy des différents modèles hybrides (attaque ICMPREDIRECTION)

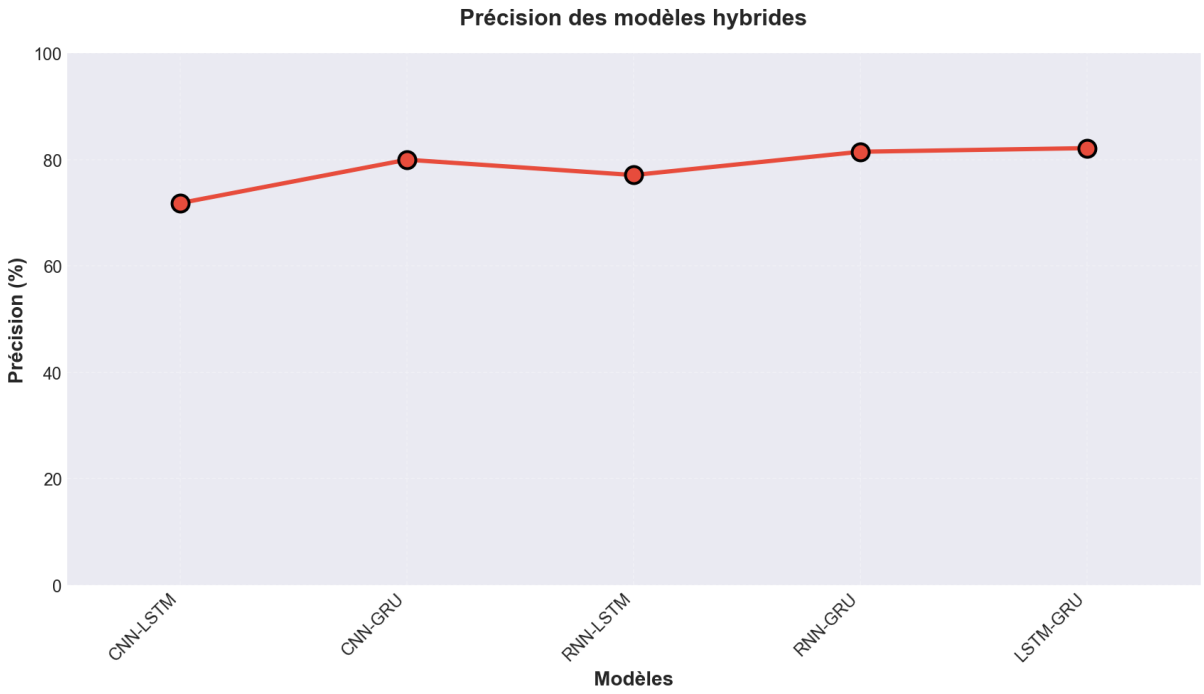


Figure III.25: Précision des différents modèles hybrides (attaque POISONING)

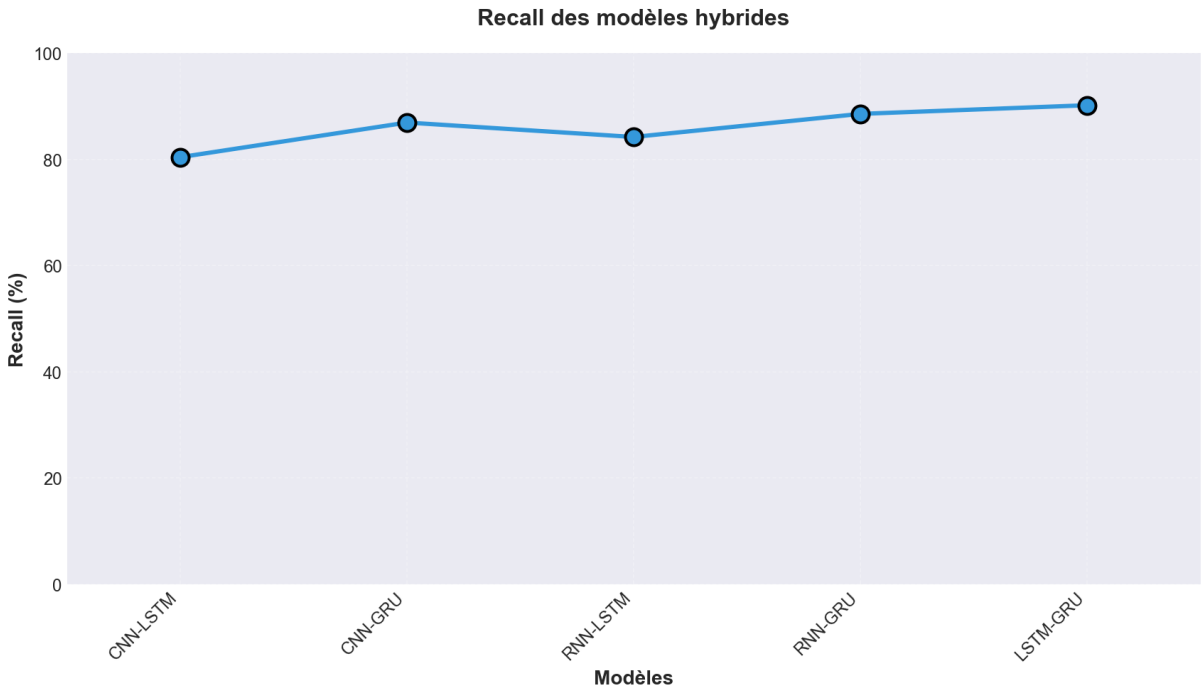


Figure III.26: Recall des différents modèles hybrides (attaque POISONING)

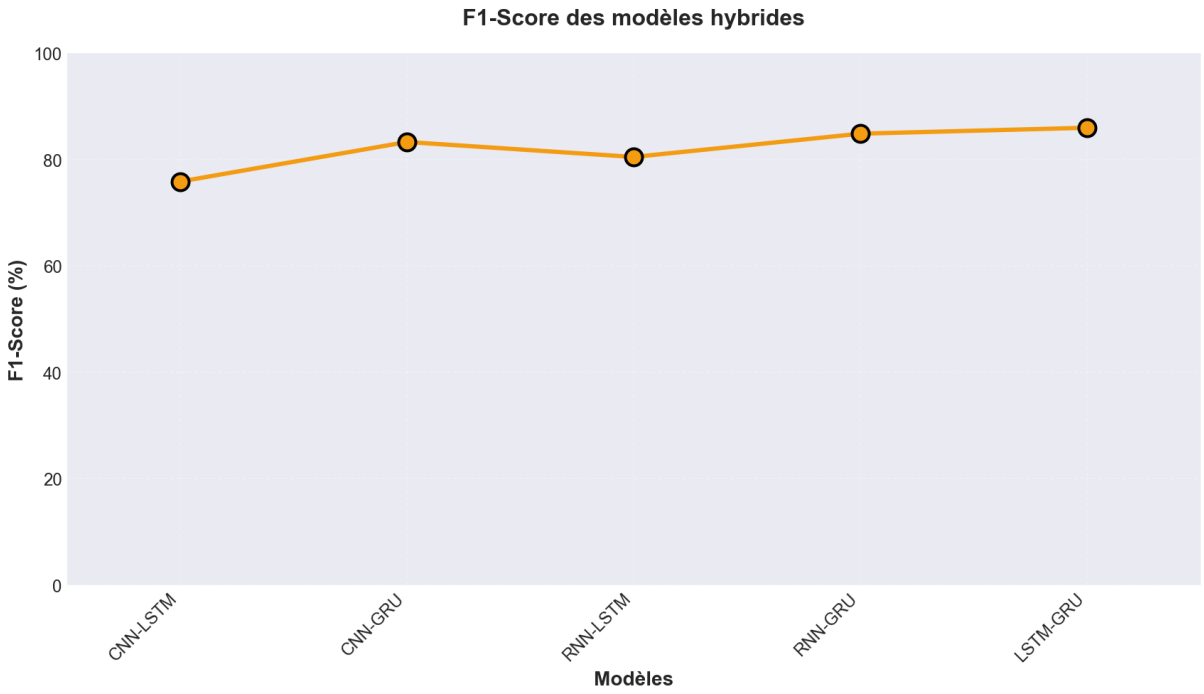


Figure III.27: F1-score des différents modèles hybrides (attaque POISONING)

L'évaluation des modèles pour la détection des attaques ARP POISONING démontre que LSTM-GRU obtient les meilleures performances avec un Recall de 90,22% (figure III.25) et une précision de 82,18% (figure III.24), confirmant sa supériorité sur toutes les architectures. La détection des attaques ARP POISONING est globalement satisfaisante pour tous les modèles, avec des Recall allant de 80,43% (CNN-LSTM) à 90,22% (LSTM-GRU), ces

performances s'expliquent par un déséquilibre des classes modéré (20,02% d'attaques), bien plus équilibré que les cas PORSCAN (3,31%) et ICMPREDIRECT (3,08%). Les fausses alertes varient de 36 à 58 selon les modèles (figure III.22), représentant un taux de faux positifs compris entre 4,90% et 7,89% des BENIGN, avec LSTM-GRU produisant seulement 36 FP, soit le meilleur résultat parmi tous les modèles. La précision (figure III.24) pour la classe ARP POISONING oscille entre 71,84% et 82,18%, ces valeurs élevées s'expliquent par un déséquilibre favorable (20% d'attaques), contrairement aux cas précédents où la précision était inférieure à 65%. Le déséquilibre modéré (79,98% BENIGN contre 20,02% ARP) constitue le facteur clé expliquant les performances robustes de tous les modèles sur cette attaque, similaires au cas TEARDROP (29,60% d'attaques) et nettement supérieures aux cas PORSCAN et ICMPREDIRECT où le déséquilibre extrême limitait la détection. LSTM-GRU confirme sa position de meilleur modèle quelle que soit la distribution des classes.

III.6. Conclusion

Dans ce chapitre, nous avons examiné l'utilité du Deep Learning pour la classification des cyber-attaques de types PORSCAN, TEARDROP, ICMPREDIRECT et ARP. Précisément, nous avons introduit une méthodologie de classification basée sur cinq architectures hybrides : LSTM-GRU, RNN-GRU, CNN-GRU, RNN-LSTM et CNN-LSTM.

Pour valider cette approche, une analyse des performances en termes d'Accuracy, Précision, Rappel et F1-score a été menée pour la classification binaire (BENIGN vs Attaque), en utilisant la base de données ICSNAD composée de quatre scénarios distincts : PORSCAN (3,31% d'attaques), TEARDROP (29,60%), ICMPREDIRECT (3,08%) et ARP POISONING (20,02%). Les résultats montrent des performances prometteuses qui dépendent de l'architecture du modèle et du déséquilibre des classes. Le modèle LSTM-GRU est le plus performant sur toutes les attaques, atteignant un rappel de 97,34% pour TEARDROP et 90,22% pour ARP, avec une précision maximale de 98% pour TEARDROP. Cependant, les modèles peinent face aux déséquilibres extrêmes : rappel maximal de 75% pour PORSCAN et 83,33% pour ICMPREDIRECT. Les architectures CNN-LSTM et RNN-LSTM sont les moins performantes, avec des rappels descendant à 33,33% et 41,67% pour PORSCAN. En revanche, pour un déséquilibre modéré (ARP et TEARDROP), tous les modèles atteignent des rappels supérieurs à 80%.

Conclusion Générale

Conclusion Générale

Les cyber-attaques contre les systèmes industriels (IC) sont en forte hausse ces dernières années et représentent aujourd'hui un risque réel qui menace les réseaux IC, les applicatifs et les systèmes d'information d de différents secteurs. En d'autres termes, la découverte périodique et permanente de ces attaques, en temps opportun, peut contribuer à les réduire en prenant les moyens de protection nécessaires. Cela nous a dirigés, dans ce mémoire, vers le développement d'un modèle de système de détection d'intrusions comme moyen d'identification des cyber-attaques, dans le but d'éviter leurs dommages, éventuellement catastrophiques.

Pour ce faire, nous avons exploré l'utilité des modèles hybrides d'apprentissage profond, à savoir LSTM-GRU, RNN-GRU, CNN-GRU, RNN-LSTM et CNN-LSTM. Ces derniers sont particulièrement utiles pour la simulation de problèmes difficiles à décrire avec des modèles physiques et mathématiques en raison de leur capacité à apprendre à partir de données d'apprentissage. Nous avons utilisé la base de données réelle ICSNAD, qui contient des flux réseau normaux et malveillants, permettant une évaluation réaliste des performances des modèles proposés.

L'évaluation comparative des modèles a été réalisée sur quatre types d'attaques distincts : PORSCAN (3,31% des échantillons), TEARDROP (29,60%), ICMPREDIRECT (3,08%) et ARP POISONING (20,02%). Les métriques utilisées sont l'accuracy, la précision, le rappel et le F1-score.

Les résultats obtenus démontrent une supériorité claire du modèle LSTM-GRU sur l'ensemble des architectures et des types d'attaques. Pour PORSCAN, LSTM-GRU atteint un rappel de 75% et une précision de 64,29% avec une accuracy de 97,80%. Pour TEARDROP, les performances sont significatives avec un rappel de 97,34% et une précision de 98%. Sur ICMPREDIRECT, malgré un déséquilibre extrême (3,08% d'attaques), LSTM-GRU parvient à détecter 83,33% des attaques. Enfin, pour ARP POISONING, le modèle obtient un rappel de 90,22% et une précision de 82,18%. Le classement général des modèles place LSTM-GRU en première position, suivi de RNN-GRU, CNN-GRU, RNN-LSTM et enfin CNN-LSTM.

La constatation la plus marquante est l'impact crucial du déséquilibre des classes sur les performances. Lorsque le pourcentage d'attaques est faible (3,08% pour ICMPREDIRECT et 3,31% pour PORSCAN), la précision chute considérablement malgré une accuracy élevée. En revanche, pour un déséquilibre modéré (20% ARP et 29,60% TEARDROP), toutes les

métriques s'améliorent significativement, avec des précisions dépassant 80% et des rappels supérieurs à 90%.

Ce travail ouvre plusieurs axes d'amélioration. Premièrement, l'utilisation de techniques de rééquilibrage des classes telles que SMOTE, ADASYN ou la pondération des fonctions de coût pourrait atténuer l'impact du déséquilibre extrême. Deuxièmement, l'exploration d'architectures plus récentes comme les Transformers ou les réseaux à attention pourrait améliorer la détection des attaques rares. Troisièmement, l'extension à la multi-classification pour distinguer les différents types d'attaques serait bénéfique.

Références Bibliographiques

Références Bibliographiques

[1] Daousis, S., Peladarinos, N., Cheimaras, V., Papageorgas, P., Piromalis, D. D., & Munteanu, R. A. (2024). Overview of Protocols and Standards for Wireless Sensor Networks in Critical Infrastructures. Future Internet, 16(1), 33. https://doi.org/10.3390/fi16010033

[2] Vogelman, S. J. (2025). The Fortnightly Review of Middle East Regional Economic & Cultural News & Developments. En ligne : https://atid-edi.com/category/fortnightly

[3] Willis, R. (2023). European Investment Bank Pledges Record Funding for Europe's Security Infrastructure, Vows More Support for Ukraine. EIB Press. En ligne : https://www.eib.org/en/press/all/2023-227-eib-pledges-record-funding-for-europe-s-security-infrastructure-vows-more-support-for-ukraine

[4] Martens, F., & Siemens, S. K. (2023). Siemens to Invest More than US$500 million in U.S. Manufacturing for Critical Infrastructure in 2023. Siemens Global Press. En ligne : https://press.siemens.com/global/en/pressrelease/siemens-invest-more-us500-million-us-manufacturing-critical-infrastructure-2023

[5] Moteff, J., & Parfomak, P. (2004). Critical Infrastructure Identification, Prioritization, and Protection. Congressional Research Service, Defense Technical Information Center. En ligne : https://apps.dtic.mil/sti/citations/ADA454016

[6] Theoharidou, M., Kandias, M., & Gritzalis, D. (2012). Securing Transportation-Critical Infrastructures: Trends and Perspectives. Global Security, Safety and Sustainability & e-Democracy, Lecture Notes of the Institute for Computer Sciences, Springer, Berlin, Heidelberg, 99, 162-171. https://doi.org/10.1007/978-3-642-33448-1_24

[7] Kumbhar, A., Koohifar, F., Guvenc, I., & Mueller, B. (2017). A survey on legacy and emerging technologies for public safety communications. IEEE Communications Surveys & Tutorials, 19(1), 97-124.

[8] Altaher, A. (2021). The Role of Modern Substation Automation Systems in Smart Grid Evolution. IEEE Smart Grid Bulletin.

[9] Zurawski, R. (2014). Industrial Communication Technology Handbook (2e éd.). CRC Press.

[10] Riedel, O., Lechler, A., & Verl, A. W. (2023). Control Architecture for Automation. Dans S. Y. Nof (Ed.), Springer Handbook of Automation. Springer, Cham. https://doi.org/10.1007/978-3-030-96729-1_16

[11] Pliatsios, D., Sarigiannidis, P., Lagkas, T., & Sarigiannidis, A. G. (2020). A Survey on SCADA Systems: Secure Protocols, Incidents, Threats and Tactics. IEEE Communications Surveys & Tutorials, 22(3), 1942-1976. https://doi.org/10.1109/COMST.2020.2987688

[12] Agbesi, J. S., Otchill, A. N., Tay, R. H., & Bamfo, N. K. (2026). Machine learning for network intrusion detection in usa critical infrastructure: challenges and opportunities. International Journal of Network Security & Its Applications (IJNSA), 18(1).

[13] Habib, M. K., & Chimsom, C. (2022). CPS: Role, characteristics, architectures and future potentials. Procedia Computer Science, 200, 1347-1358.

[14] Modbus Organization. (2012). Modbus Application Protocol Specification V1.1b3. Modbus DA.

[15] Kullaa, L. (2026). Communication Performance Improvement of PROFINET Networks in Industrial Crane Systems (Master's thesis). Aalto University.

[16] Marshall, P. S., & Rinaldi, J. S. (2016). Industrial Ethernet: How to Plan, Install, and Maintain TCP/IP Ethernet Networks (3e éd.). International Society of Automation / John Wiley & Sons.

[17] Zaid, B. (2014). Étude d'un système de supervision et de contrôle SCADA de la région de transport est RTE Skikda (Mémoire de Master). Université de Skikda.

[18] Sridevi, G. (2018). Review Paper on Ethernet for Control Automation Technology (EtherCAT). International Journal of Engineering Research & Technology (IJERT), 6(13).

[19] Schneider Electric. (2024). TeSys™ T LTMR Motor Management Controller CANopen Communication Guide (Technical Report).

[20] Lian, F. L., Moyne, J. R., & Tilbury, D. M. (2001). Performance evaluation of control networks: Ethernet, ControlNet, and DeviceNet. IEEE Control Systems Magazine, 21(1), 66-83.

[21] Mir, M. T. (2024). Security analysis of Message Queuing Telemetry Transport (MQTT) (PhD Thesis). Universitat Politècnica de Catalunya.

[22] Center for Strategic & International Studies (CSIS). (2026). Significant Cyber Incidents. En ligne : https://www.csis.org/

[23] JD Supra. (2021). Ransomware on the Rise in Critical Infrastructure Sector. En ligne : https://www.jdsupra.com/legalnews/ransomware-on-the-rise-in-critical-1687319/

[24] Oz, H., Aris, A., Levi, A., & Uluagac, A. S. (2022). A Survey on Ransomware: Evolution, Taxonomy, and Defense Solutions. ACM Computing Surveys, 54(11), Article 238.

[25] Kumar, S., Kumar, H., & Gunnam, G. R. (2019). Security Integrity of Data Collection from Smart Electric Meter under a Cyber Attack. Dans Proceedings of the 2019 2nd International Conference on Data Intelligence and Security (ICDIS), Texas, USA, pp. 9–13.

[26] Psiaki, M. L., & Humphreys, T. E. (2016). GNSS Spoofing and Detection. Proceedings of the IEEE, 104(6), 1258–1270.

[27] Schuckers, S. A. (2002). Spoofing and anti-spoofing measures. Information Security Technical Report, 7(4), 56–62.

[28] Radware. (2025). ARP Poisoning - DDoS Knowledge Center. En ligne : https://www.radware.com/security/ddos-knowledge-center/ddospedia/arp-poisoning/

[29] Conti, M., Dragoni, N., & Lesyk, V. (2016). A survey of man in the middle attacks. IEEE Communications Surveys & Tutorials, 18(3), 2027–2051.

[30] Verizon. (2019). Data Breach Investigations Report. En ligne : https://www.verizon.com/business/resources/reports/dbir/

[31] Tufail, S., Batool, S., & Sarwat, A. I. (2021). False data injection impact analysis in AI-based smart grid. Dans Proceedings of the IEEE SoutheastCon 2021, Atlanta, GA, USA, pp. 1–7.

[32] Pratama, A., & Rafrastara, F. A. (2012). Computer worm classification. International Journal of Computer Science and Information Security, 10(1), 21.

[33] Riggs, H., Tufail, S., Parvez, I., Tariq, M., Khan, M. A., Amir, A., Vuda, K. V., & Sarwat, A. I. (2023). Impact, Vulnerabilities, and Mitigation Strategies for Cyber-Secure Critical Infrastructure. Sensors, 23(8), 4060. https://doi.org/10.3390/s23084060

[34] Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep machine learning-a new frontier in artificial intelligence research. IEEE Computational Intelligence Magazine, 5(4), 13-18.

[35] Setiowati, S., Zulfanahri, E. F., & Ardiyanto, I. (2017). A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods. Dans Proceedings of the International Conference on Information Technology (ICIT).

[36] Mimoune, Z., & Ouahab, A. (2019). Développement d'une Architecture Basée sur l'Apprentissage Profond (Deep Learning) pour la Détection d'Intrusion dans les Réseaux (Mémoire de fin d'études). Université Ahmed Draia – Adrar.

[37] Moualek, D. Y. (2017). Deep Learning pour la classification des images (Mémoire de Master). Université Abou Bakr Belkaid – Tlemcen.

[38] Gers, F., Schraudolph, N., & Schmidhuber, J. (2002). Learning Precise Timing with LSTM Recurrent Networks. Journal of Machine Learning Research, 3, 115-143.

[39] Rivals, I. (1995). Modélisation et commande par réseaux de neurones : application au pilotage d'un véhicule autonome (Thèse de Doctorat). Université Pierre et Marie Curie – Paris VI.

[40] Zhou, X., Cheng, Z., Wang, C., et al. (2026). A dataset collected in real-world industrial control systems for network attack detection. Scientific Data (Nature), 13(399). https://doi.org/10.1038/s41597-026-06738-x

[41] Labonne, M., Olivereau, A., & Zeghlache, D. (2018). Automatisation du processus d'entraînement d'un ensemble d'algorithmes de machine learning optimisés pour la détection d'intrusion. Rapport de recherche académique / Conférence.