République Algérienne Démocratique et Populaire
Ministére de l'Enseignement Supérieur et de la Recherche Scientifique
Université A. Mira de Béjaïa
Faculté des Sciences Exactes
Département d'Informatique

Mémoire de Fin de Cycle

En vue de l'obtention du diplôme de Master Recherche en Informatique

Option :

Réseau et sécurité

Thème

Sécurité des Systèmes de Transport Ferroviaire Intelligents

Réalisé par :

BEGRICHE KATIA

Devant le jury composé de :

Présidente :	Mme KESSIRA	Dalila	Université de Béjaïa
Encadrante :	Mme ZAMOUCHE	Djamila	Université de Béjaïa
Co-encadrante :	Mme YESSAD	Nawel	Université de Béjaïa
Examinatrice :	Mme CHERIFI	Feriel	Université de Béjaïa
Examinatrice :	Mme MECHIOURI	Sara	Université de Béjaïa

Promotion 2023/2024

Remerciements

Avant tout, je remercie le bon Dieu qui m'a accordé la santé et le savoir pour accomplir ce travail.

Comme je tiens à remercier mes très chers parents pour leurs sacrifices, leurs patiences et leurs soutiens moral.

Sans l'aide précieuse, le soutien, la compétence et la relecture critique de mon encadrante, Mme Djamila Zamouche, ainsi que de Mme Nawel YESSAD, ce travail n'aurait pas pu être accompli. Leur apport précieux, tant sur le fond que sur la forme de ce mémoire, m'a orienté vers la connaissance.

Je tiens également à exprimer ma gratitude envers les membres du jury qui m'ont fait l'honneur d'examiner et de juger mon mémoire.

Dedicace

Je dédie ce modeste travail et exprime ma profonde gratitude :

À ma chère, douce et tendre mère,
Tout ce que je peux t'offrir ne pourra jamais exprimer pleinement mon respect, mon amour éternel et ma reconnaissance pour les sacrifices que tu as consentis pour mon instruction et l'éducation que tu m'as prodiguée.

À mon cher père,
qui a toujours fait preuve de maturité, qui maintient toujours courage et calme face à l'adversité, dont la force étonne et la résilience inspire.

Merci pour ton amour, ton soutien et tes sacrifices.

À ma chère sœur et à mon frère,

À tous mes amis et collègues,

À tous ceux qui m'ont aidé,

À tous ceux qui me sont chers,

À tous ceux que j'ai omis.

Avec toute ma gratitude.

Table des Matières

Table des figures

Liste des tableaux

Liste des Abréviations

ATO le Controle Automatique des trains

ATP Système de Protection des trains

ATS la Supervision Automatique des trains

CBTC Communication Based Tran Control

DoS attaque par déni de service

DS-SPS semi-persistent scheduling basé sur la détection

ECC-CLPS Elliptic Cuve Cryptography-Based Certificateless Proxy Signature

FHSS Frequency Hopping Spread Spectrum

G-ZC Gound-Zone Control

HASC Hybrid Approach for Secure Communications

HMM Modèle de Markov caché

IoT Internet des Objet

LSTM Long Short-Term Memory

LTE Long Term Evolution

MA Autorité de Mauvement

ML Machine Learning

On-ZC Onboard Zone Controllers

PIS Système d'Information des Passagers

RNN Recurrent Neural Network

SCC Système de Communication et de Contrôle

SDR Software Defined Radio

SINR Signal-to-Interface Noise Radio

STFI Systèmes de Transport Ferroviaire Intelligents

SVM Support Vector Machine

T2T Train-à-Train

T2W train-terre

TC-CBTC Train-Centric Communication-Based Train Control

TMS Système de Gestion du Trafic Ferroviaire

Introduction générale

Les systèmes de transport ferroviaire intelligents (STFI) représentent une avancée majeure dans l'optimisation et la sécurisation des réseaux ferroviaires modernes. Ces systèmes intégrés, combinant technologies de communication, de contrôle et de gestion de l'information, sont conçus pour améliorer la performance, la sécurité et l'efficacité des opérations ferroviaires. Le TC-CBTC (Transmission-based Communication and Control) est un exemple éminent de ces technologies, permettant une gestion optimisée et sécurisée des trains grâce à des communications continues et fiables entre les trains et les centres de contrôle.

Avec l'augmentation de la complexité et de l'interconnectivité des systèmes ferroviaires, la sécurité des communications devient une préoccupation majeure. La principale problématique de ce mémoire est de sécuriser la communication entre les trains dans un système TC-CBTC, en garantissant l'intégrité et la confidentialité des informations échangées. Les cyberattaques et les anomalies peuvent compromettre non seulement la sécurité des passagers, mais aussi l'efficacité opérationnelle et la fiabilité des services ferroviaires.

Plusieurs solutions ont été proposées pour sécuriser ces communications. Les méthodes de cryptographie traditionnelle utilisent des techniques de chiffrement pour protéger les données en transit, mais elles peuvent être vulnérables aux attaques de type "man-in-the-middle" et nécessitent une gestion complexe des clés. La détection d'anomalies basée sur des règles s'appuie sur des règles prédéfinies pour identifier les comportements anormaux, mais elle est limitée par son incapacité à détecter des attaques inconnues ou évolutives. Enfin, l'apprentissage automatique pour la sécurité utilise des modèles d'apprentissage automatique pour identifier des schémas anormaux dans les données de communication. Bien que prometteurs, ces systèmes peuvent générer des faux positifs et nécessitent des données d'entraînement diversifiées et conséquentes. Les limites de ces solutions comprennent une capacité restreinte à détecter de nouvelles menaces, une gestion et une intégration des systèmes complexes, ainsi qu'une possible surcharge computationnelle.

Ce mémoire propose une approche hybride alliant apprentissage automatique et cryptographie pour surmonter les limitations actuelles. Nous utilisons des réseaux de neurones récurrents (RNN) pour détecter de manière proactive les anomalies dans les communications. Ensuite, une méthode d'isolation forest affine ces détections pour plus de précision. Enfin, une machine à vecteurs de support (SVM) effectue la classification finale des com-

portements anormaux. Cette architecture en cascade permet de combiner la robustesse des techniques cryptographiques avec la capacité des modèles d'apprentissage automatique à identifier les comportements anormaux et les menaces évolutives.

Notre travail est structuré en quatre chapitres.

Dans le chapitre I , nous avons introduit les systèmes de transport ferroviaire intelligents (STFI) . Nous avons commencé par définir les principaux composants des STFI, en mettant en lumière leur rôle crucial dans l'amélioration de l'efficacité opérationnelle et de la sécurité. Ensuite, nous avons examiné comment les avancées technologiques comme l'intelligence artificielle (IA) sont intégrées pour optimiser la gestion des réseaux ferroviaires, notamment en matière de planification des horaires, de prédiction des pannes et de gestion des actifs.

Le chapitre II explore de manière approfondie l'état de la sécurité dans les systèmes de transport ferroviaire intelligent, mettant en lumière les avancées technologiques récentes, les défis persistants et les solutions innovantes. Il aborde diverses méthodes telles que l'utilisation de techniques de brouillage, les algorithmes d'apprentissage automatique pour la détection d'intrusions, les modèles de Markov cachés pour la fusion d'informations réseau et dispositif, ainsi que les architectures sécurisées utilisant des modules matériels de sécurité et des mécanismes de chiffrement avancés.

Le chapitre III présente notre approche pour la sécurisation des communication dans le système TC-CBTC en utilisant des techniques d'apprentissage automatiques et des mécanismes de cryptographie.

Enfin, le chapitre IV évalue la performance de notre modèle en mesurant la précision, le rappel, l'exactitude et la F1-mesure.

1

Systèmes de Transport Ferroviaire Intelligents

1.1 Introduction

Le transport ferroviaire évolue rapidement grâce à l'intégration de technologies avancées qui visent à améliorer la sécurité, l'efficacité et la fiabilité des opérations. Ce chapitre présente les Systèmes de Transport Ferroviaire Intelligents (STFI), en mettant l'accent sur les systèmes Communication-Based Train Control (CBTC) et Train-Centric Communication-Based Train Control (TC-CBTC). Nous explorerons leurs architectures, leurs infrastructures, ainsi que les enjeux de sécurité associés, tout en soulignant l'impact de l'Intelligence Artificielle (IA) dans ce domaine.

1.2 Systèmes de Transport Ferroviaire Intelligents (STFI)

1.2.1 Définition

Les STFI sont des ensembles intégrés de technologies de communication, de contrôle et de gestion des informations, visant à améliorer la performance, la sécurité et l'efficacité des opérations ferroviaires. Ces systèmes incluent des solutions pour la gestion du trafic, la maintenance prédictive, la surveillance de la sécurité et l'information des passagers,

permettant une optimisation continue des réseaux ferroviaires [1].

1.2.2 Composants Clés des STFI

Les STFI intègrent divers composants technologiques avancés pour améliorer l'efficacité, la sécurité et la fiabilité des opérations ferroviaires.

1.2.2.1 Systèmes de Communication et de Contrôle (SCC)

Les SCC sont essentiels pour assurer une interaction fluide entre les trains, les infrastructures au sol, et les centres de contrôle. Ces systèmes utilisent des technologies comme la Communication-Based Train Control (CBTC) et les communications de train à train (T2T).

1.2.2.2 Systèmes de Gestion du Trafic Ferroviaire (TMS)

Les TMS optimisent le flux des trains, la planification des horaires et la coordination des mouvements pour minimiser les retards et éviter les conflits.

— Optimisation des Horaires : Les TMS utilisent des algorithmes sophistiqués pour planifier les horaires des trains, en prenant en compte divers facteurs comme les horaires de pointe, les travaux de maintenance et les incidents imprévus [2].

— Importance de l'Optimisation des Horaires : Une optimisation efficace des horaires réduit les conflits entre trains et les temps d'attente aux intersections, ce qui diminue les retards et augmente la capacité de la ligne en réduisant les intervalles entre les trains. De plus, des horaires fiables et réguliers améliorent la satisfaction des passagers en réduisant les temps d'attente et en offrant des correspondances plus fluides. Les opérateurs ferroviaires bénéficient également de l'optimisation des horaires, car elle permet une gestion plus efficace des ressources, y compris les trains, le personnel et l'énergie. Les systèmes de gestion du trafic (TMS) jouent un rôle clé en intégrant des données en temps réel pour ajuster dynamiquement les horaires et gérer les perturbations, assurant ainsi une gestion optimale du trafic ferroviaire [2].

1.2.2.3 Surveillance et Maintenance Prédictive

La surveillance et la maintenance prédictive utilisent des capteurs et des technologies de l'Internet des Objets (IoT) pour surveiller en temps réel l'état des infrastructures et des trains, permettant une maintenance proactive ; les capteurs IoT collectent des données sur

l'état des équipements ferroviaires, permettant de détecter les anomalies avant qu'elles ne provoquent des pannes [3].

1.2.2.4 Systèmes de Sécurité

Les systèmes de sécurité ferroviaire comprennent des technologies comme les systèmes de protection automatique des trains (ATP), la surveillance vidéo et les systèmes de détection d'intrusion pour améliorer la sécurité des passagers et du personnel [4].

1.2.2.5 Systèmes d'Information des Passagers (PIS)

Les PIS fournissent des informations en temps réel concernant les horaires, les retards, les correspondances et les services à bord, améliorant ainsi l'expérience des usagers ; Les PIS utilisent des écrans numériques et des applications mobiles pour informer les passagers en temps réel, ce qui améliore la satisfaction des usagers et réduit l'incertitude lors des voyages [5].

1.3 Caractéristiques des Systèmes de Transport Ferroviaire Intelligent (STFI)

Les STFI se distinguent par leur utilisation de technologies de pointe telles que les capteurs, les communications sans fil et l'intelligence artificielle pour surveiller et contrôler les opérations ferroviaires en temps réel. Ces systèmes sont conçus pour être interopérables, intégrant différents sous-systèmes ferroviaires pour une gestion coordonnée et une optimisation globale du réseau. De plus, ils offrent des services personnalisés aux passagers grâce à des applications mobiles conviviales et des informations en temps réel sur les horaires et les correspondances. Enfin, les STFI intègrent des fonctionnalités avancées de sécurité pour assurer la sûreté des opérations et la sécurité des passagers et du personnel [5].

1.4 Le Contrôle de Train Basé sur la Communication (CBTC)

1.4.1 Définition de CBTC

Le Contrôle de Train Basé sur la Communication (CBTC) est un système avancé de signalisation ferroviaire qui utilise des communications bidirectionnelles en temps réel entre les trains et les équipements au sol pour assurer le contrôle de la circulation ferroviaire. Contrairement aux systèmes traditionnels qui reposent sur des signaux fixes, le CBTC permet une gestion dynamique et précise des trains, augmentant ainsi la capacité et la sécurité des réseaux ferroviaires urbains [6].

1.4.2 Fonctionnement et Principes du CBTC

Le CBTC permet une gestion plus précise et dynamique de la circulation ferroviaire. Les trains équipés de CBTC transmettent en continu leur position, leur vitesse et d'autres données opérationnelles à un centre de contrôle via des communications sans fil. Ce système utilise ces données pour calculer des commandes optimisées pour chaque train, ajustant leur vitesse et leur espacement en temps réel pour maximiser la sécurité et l'efficacité. Le CBTC permet une réduction significative des distances de sécurité entre les trains, augmentant ainsi la capacité du réseau ferroviaire. Il offre également des fonctionnalités avancées telles que le contrôle automatique des trains (ATO), la supervision automatique des trains (ATS), et la protection automatique des trains (ATP). Le CBTC optimise l'utilisation de l'infrastructure ferroviaire en permettant un contrôle plus fin et adaptatif des trains, ce qui se traduit par une augmentation de la capacité, une amélioration de la sécurité et une réduction des coûts d'exploitation [6].

1.4.3 Avantages du CBTC

Le Contrôle de Train Basé sur la Communication (CBTC) présente plusieurs avantages significatifs par rapport aux systèmes traditionnels de signalisation ferroviaire, ce qui explique son adoption croissante dans le monde entier.

1. Amélioration de la sécurité : Le CBTC utilise des communications en temps réel pour surveiller en continu la position et la vitesse des trains, ce qui permet d'éviter les collisions et de réduire les risques d'accidents. Les systèmes de protection automatique des trains (ATP) intégrés au CBTC garantissent que les trains respectent les limitations de vitesse et les distances de sécurité.
2. Augmentation de la capacité : En permettant un suivi plus précis et une gestion plus dynamique des trains, le CBTC réduit les distances de sécurité nécessaires

entre les trains. Cela permet d'augmenter le nombre de trains pouvant circuler sur une même ligne, améliorant ainsi la capacité globale du réseau ferroviaire.

3. Efficacité opérationnelle : Le CBTC optimise les horaires des trains et réduit les temps d'attente aux stations, ce qui améliore l'efficacité opérationnelle. Les trains peuvent circuler plus fréquemment et de manière plus fiable, ce qui réduit les temps de trajet pour les passagers.
4. Flexibilité et adaptabilité : Le CBTC peut facilement s'adapter aux changements dans les conditions de trafic et aux perturbations. En cas de panne ou de travaux sur la voie, le système peut réajuster rapidement les horaires et les itinéraires des trains pour minimiser les interruptions de service.
5. Réduction des coûts d'exploitation : En automatisant de nombreuses fonctions de contrôle et de supervision des trains, le CBTC réduit les besoins en personnel et en maintenance, ce qui diminue les coûts d'exploitation à long terme.

1.4.4 Applications du CBTC

1. Transports urbains : Le CBTC est largement utilisé dans les systèmes de métro et de transport urbain rapide. Des villes comme New York, Londres, Paris et Hong Kong ont adopté le CBTC pour améliorer la capacité et la fiabilité de leurs réseaux de métro.
2. Transport ferroviaire interurbain : Le CBTC est également appliqué dans les réseaux de trains interurbains pour augmenter la capacité et améliorer la gestion du trafic sur les lignes à haute densité. Cela est particulièrement important dans les corridors de transport très fréquentés.

1.4.5 Technologies et Composants des Systèmes de Contrôle de Train Basé sur la Communication (CBTC)

Le système de Contrôle de Train Basé sur la Communication (CBTC) repose sur un ensemble de technologies et de composants clés qui permettent une surveillance et un contrôle précis et en temps réel des mouvements des trains. Ces technologies et composants incluent des équipements embarqués, des infrastructures au sol, et des systèmes de communication sophistiqués.

1.4.5.1 Équipements Embarqués

Les trains équipés de CBTC possèdent plusieurs dispositifs essentiels qui permettent le suivi et le contrôle de leurs mouvements :

— Unités de contrôle de bord (Onboard Controllers) : Ces unités sont responsables du traitement des données de position et de vitesse du train. Elles reçoivent des commandes du centre de contrôle et exécutent des actions comme ajuster la vitesse ou arrêter le train si nécessaire.
— Capteurs de position : Ils déterminent la position exacte du train en temps réel. Les capteurs GPS, les odomètres et les balises installées le long des voies sont souvent utilisés pour cette tâche.
— Systèmes de communication embarqués : Ils assurent la transmission bidirectionnelle continue des données entre le train et le centre de contrôle. Cela inclut les radios numériques et les réseaux sans fil dédiés.

la figure 1.1 représente l'architecture de CBTC [7].

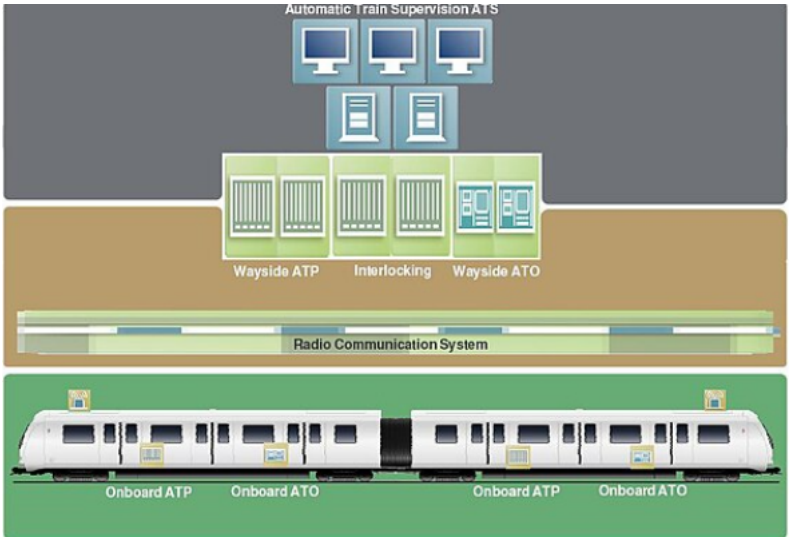

FIGURE 1.1 – L'architecture d'un système CBTC.

1.4.5.2 Infrastructures au Sol

L'infrastructure au sol du CBTC est composée de divers éléments qui facilitent la gestion et le contrôle des trains :

— Balises et antennes : Installées le long des voies, elles communiquent avec les trains pour fournir des données de position supplémentaires et assurer une redondance dans le système de localisation.
— Centres de contrôle : Équipés de systèmes informatiques puissants, ces centres reçoivent les données des trains, calculent les commandes optimales et les renvoient aux trains. Ils jouent un rôle crucial dans la gestion du trafic ferroviaire.
— Réseaux de communication au sol : Ces réseaux relient les différentes composantes du système CBTC, assurant une transmission rapide et fiable des données. Ils peuvent inclure des réseaux câblés et sans fil, selon les besoins et les contraintes du réseau ferroviaire.

1.4.5.3 Systèmes de Communication

Le cœur du CBTC est son système de communication sophistiqué, qui assure une interaction continue et en temps réel entre les trains et l'infrastructure au sol :

— Communication bidirectionnelle en temps réel : Utilise des technologies avancées comme le LTE (Long Term Evolution) ou des réseaux dédiés pour garantir que les informations sur la position, la vitesse et les commandes de contrôle sont transmises sans délai.

— Redondance et sécurité : Les systèmes CBTC sont conçus pour être redondants et sécurisés, utilisant plusieurs canaux de communication et des protocoles de sécurité robustes pour éviter les pannes et les interférences.

1.4.6 Train-Centric Communication-Based Train Control (TC-CBTC)

1.4.7 Définition

Le TC-CBTC est un système de contrôle de train basé sur la communication qui adopte une approche décentralisée, mettant l'accent sur les communications directes entre les trains. Le TC-CBTC utilise des contrôleurs embarqués sur chaque train, appelés Onboard-Zone Controllers (On-ZC). Ces contrôleurs embarqués gèrent le contrôle et la gestion des trains de manière autonome, permettant une communication plus rapide, une réduction des délais de transmission et une amélioration de la flexibilité opérationnelle [6].

1.4.7.1 Caractéristiques du TC-CBTC

Les principales caractéristiques du TC-CBTC incluent :

1. **Architecture Décentralisée :** Chaque train est équipé de son propre contrôleur embarqué, éliminant ainsi le besoin de contrôleurs de zone au sol. Cette décentralisation permet aux trains de fonctionner de manière plus autonome.
2. **Communication Train-à-Train (T2T) :** Les trains communiquent directement entre eux, ce qui élimine la dépendance aux infrastructures au sol pour certaines fonctions critiques et permet une transmission d'information plus rapide.
3. **Optimisation des Ressources :** Le TC-CBTC utilise des schémas de répartition des ressources tels que le semi-persistent scheduling basé sur la détection (DS-SPS), permettant une gestion optimisée des ressources de communication entre les trains.

La figure 1.2 représente l'architecture du système TC-CBTC [8].

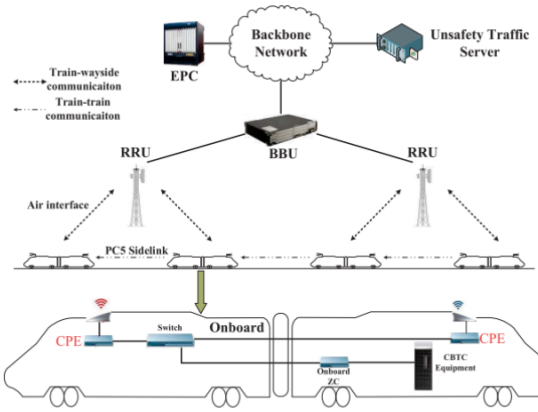

FIGURE 1.2 – L'architecture du système TC-CBTC.

1.4.7.2 Autorité de Mouvement (MA)

Dans le cadre du TC-CBTC, l'autorité de mouvement (MA) est une directive émise par le système de contrôle de train qui autorise un train à se déplacer jusqu'à un point spécifique de la voie. Avec l'architecture décentralisée du TC-CBTC, l'On-ZC de chaque train joue un rôle clé dans la gestion de l'autorité de mouvement. Le MA peut être transmis directement entre les trains grâce à la communication T2T, ce qui permet une mise à jour rapide et en temps réel des autorisations de mouvement. Cette approche permet de réduire les délais et d'améliorer la sécurité opérationnelle en assurant que chaque train possède des informations précises et à jour sur ses autorisations de mouvement.

1.4.8 Différence entre CBTC et TC-CBTC

Le CBTC (Communication-Based Train Control) et le TC-CBTC (Train-Centric Communication-Based Train Control) sont deux systèmes de contrôle de train basés sur la communication, mais ils présentent des différences fondamentales en termes de conception, de fonctionnement et d'architecture.

Bien que les deux systèmes visent à améliorer la sécurité et l'efficacité des opérations ferroviaires, le TCCBTC offre une architecture plus décentralisée et flexible, permettant une gestion plus autonome et directe des trains.

Aspect	CBTC	TCCBTC
Architecture	Centralisée	Décentralisée
Communication	Train-Terre (T2W)	Train-à-Train (T2T)
Contrôleur	Ground-Zone Controller (G-ZC)	Onboard-Zone Controller (On-ZC)
Flexibilité	Moins flexible en cas de pannes locales	Plus flexible et résilient
Complexité des équipements	Plus élevée au sol	Réduite au sol, augmentée à bord
Optimisation des ressources	Dépendante du centre de contrôle	Gérée localement par chaque train

TABLE 1.1 – Tableau comparatif de CBTC et TCBTC

1.5 Enjeux dans les systèmes ferroviaires

1.5.1 Définition

Les enjeux dans les systèmes ferroviaires concernent toutes les mesures et les précautions nécessaires pour garantir la sûreté et la fiabilité des opérations ferroviaires. Cela inclut la protection des passagers, du personnel, des infrastructures et des équipements contre divers risques tels que les accidents, les défaillances techniques, les cyberattaques, et les actes malveillants. La sécurité ferroviaire vise à prévenir les incidents et à minimiser les conséquences potentielles de ces incidents sur la vie humaine, l'environnement et l'économie.

1.5.2 Enjeux de Sécurité

1.5.2.1 Cybersécurité

La Cybersécurité dans les systèmes ferroviaires concerne la protection des systèmes de contrôle, des réseaux de communication et des données contre les cyberattaques. Cela inclut la prévention des intrusions, la détection des activités suspectes et la mise en place de mesures de réponse pour atténuer les impacts des cyber menaces. Les cyberattaques peuvent viser à perturber les opérations, à accéder à des informations sensibles ou à compromettre la sécurité des passagers et des infrastructures.

1.5.2.2 Sécurité des données

La sécurité des données et de la vie privée concerne la protection des informations collectées, traitées et stockées par les systèmes ferroviaires, y compris les données personnelles des passagers et les données opérationnelles. Cela inclut des mesures pour garantir la confidentialité, l'intégrité et la disponibilité des données, ainsi que pour se conformer aux réglementations sur la protection des données personnelles. Des protocoles de chiffrement, des contrôles d'accès stricts et des politiques de gestion des données sont essentiels pour protéger ces informations.

1.5.2.3 Authentification

L'authentification dans les systèmes ferroviaires est un enjeu de sécurité majeur, car elle garantit que seules les entités autorisées peuvent accéder et interagir avec les systèmes de contrôle ferroviaire. Cette mesure empêche les intrusions et les accès non autorisés, réduisant ainsi le risque de sabotages ou d'attaques malveillantes qui pourraient compromettre la sécurité des opérations ferroviaires. En outre, l'authentification garantit la protection de l'intégrité et de la confidentialité des données échangées en temps réel entre les trains, les centres de contrôle et les infrastructures au sol, empêchant ainsi toute manipulation ou interception de ces informations essentielles. En veillant à ce que les informations proviennent de sources fiables. L'authentification est légale et permet de coordonner précisément et de manière sécurisée les mouvements de trains, réduisant ainsi les risques d'accidents et de perturbations [9] [10].

1.5.2.4 Sécurité physique

La sécurité physique dans les systèmes ferroviaires se réfère à la protection des infrastructures ferroviaires, y compris les voies, les gares, les trains, et les équipements critiques contre les menaces physiques telles que le vandalisme, le sabotage, le terrorisme, et les intrusions non autorisées. Cela implique des mesures telles que la surveillance vidéo, les clôtures de sécurité, les systèmes de détection d'intrusion et la présence de personnel de sécurité.

1.6 Sûreté de fonctionnement

1.6.1 Défaillances matérielles et logicielles

Les défaillances matérielles et logicielles sont des pannes ou des dysfonctionnements qui peuvent survenir dans les composants physiques (comme les rails, les trains, les systèmes de signalisation) ou les systèmes informatiques et logiciels utilisés pour contrôler et gérer les opérations ferroviaires. Les défaillances matérielles peuvent inclure des ruptures de rail ou des défaillances mécaniques, tandis que les défaillances logicielles peuvent inclure des bugs, des erreurs de programmation ou des problèmes de compatibilité.

1.6.2 Facteurs humains

Les facteurs humains se réfèrent aux erreurs ou aux comportements des personnes impliquées dans l'exploitation, la maintenance et la gestion des systèmes ferroviaires qui peuvent affecter la sécurité. Cela inclut les erreurs de jugement, la fatigue, le manque de formation, et la non-conformité aux procédures de sécurité. Les facteurs humains peuvent jouer un rôle critique dans la prévention des accidents et des incidents.

1.7 L'Intelligence Artificielle (IA) dans les Systèmes de Transport Ferroviaire

L'Intelligence Artificielle (IA) dans les systèmes de transport ferroviaire désigne l'application de techniques et d'algorithmes avancés de machine learning et de traitement des données pour améliorer divers aspects de l'exploitation ferroviaire. L'IA peut être utilisée pour optimiser la gestion du trafic, prévoir les besoins en maintenance, améliorer la sécurité, et offrir des services plus personnalisés aux passagers. Grâce à l'analyse des données en temps réel et à l'apprentissage automatique, les systèmes ferroviaires peuvent devenir plus efficaces, sûrs et réactifs aux défis opérationnels [1].

1.7.1 Algorithmes

Dans cette partie, nous allons présenter les principaux algorithmes d'apprentissage automatique et leurs caractéristiques spécifiques, en mettant en lumière leur pertinence pour les systèmes de transport ferroviaire intelligents.

1. **Support Vector Machine (SVM)**
   Les machines à vecteurs de support (SVM) sont des algorithmes d'apprentissage

supervisé utilisés principalement pour la classification et la régression. L'idée principale de SVM est de trouver l'hyperplan qui sépare les classes de données de manière optimale dans un espace de caractéristiques. Les points de données les plus proches de l'hyperplan sont appelés vecteurs de support, et l'objectif est de maximiser la marge entre ces vecteurs et l'hyperplan [11].

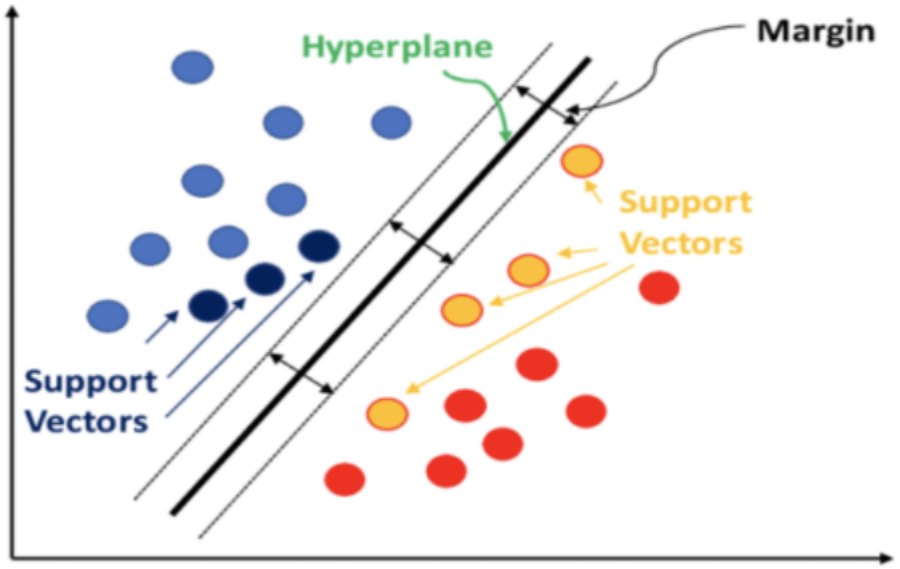

FIGURE 1.3 – Support Vector Machine (SVM) [12].

Caractéristiques :

(a) **Séparation Maximale :** SVM cherche à maximiser la distance entre les classes de données.

(b) **Kernel Trick :** Utilisation de fonctions noyaux pour traiter les données non linéaires en les projetant dans un espace de dimensions supérieures où elles deviennent séparables linéairement.

(c) **Complexité :** Adapté pour des ensembles de données de taille moyenne à grande, mais peut être coûteux en termes de calcul pour des ensembles très grands.

2. **Recurrent Neural Network (RNN)**
Les réseaux de neurones récurrents (RNN) sont une classe de réseaux de neurones adaptés pour traiter les données séquentielles ou temporelles. Contrairement aux réseaux de neurones classiques, les RNN possèdent des connexions récurrentes qui leur permettent de conserver une mémoire de l'état précédent, ce qui est crucial pour l'analyse des séries temporelles et des données séquentielles [13].

Caractéristiques :

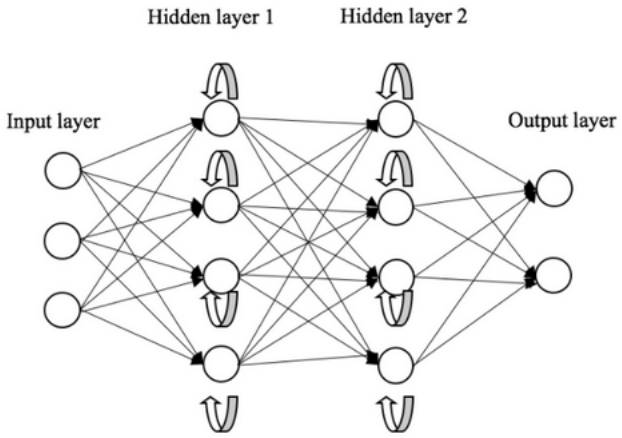

FIGURE 1.4 – Forme général d'un RNN [14].

(a) **Mémoire :** Capacité à conserver l'information des états précédents grâce aux connexions récurrentes.

(b) **Applications :** Utilisés pour la reconnaissance vocale, la traduction automatique, la prédiction de séries temporelles, etc.

(c) **Vanishing Gradient Problem :** Problème de gradients qui disparaissent ou explosent, souvent résolu par des variantes comme les LSTM (Long Short-Term Memory) et les GRU (Gated Recurrent Units).

3. **Isolation Forest**
L'Isolation Forest est un algorithme d'apprentissage automatique non supervisé utilisé principalement pour la détection des anomalies. L'idée principale est d'isoler les anomalies plutôt que de profiler le comportement normal. L'algorithme construit des arbres de manière aléatoire et mesure le chemin d'isolation des points de données, les anomalies étant celles qui nécessitent moins de partitions pour être isolées [15].

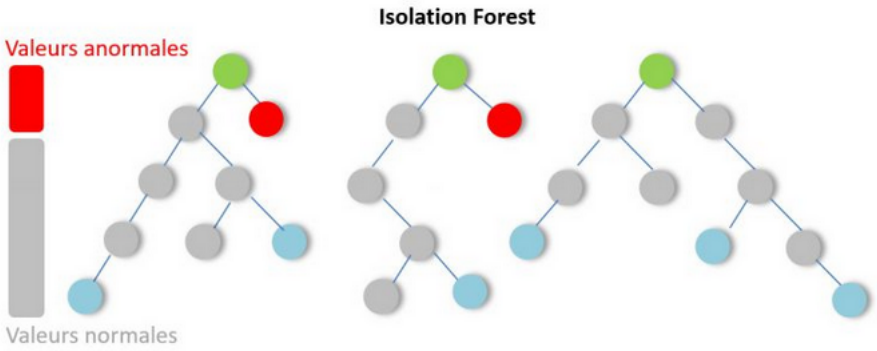

FIGURE 1.5 – Forme général de Isolation forest [16].

Caractéristiques :

(a) **Isolation :** L'isolation des anomalies est plus rapide car elles sont plus susceptibles de se trouver dans des régions rares.

(b) **Complexité :** L'algorithme est efficace en termes de calcul et évolutif pour de grands ensembles de données.

(c) **Applications :** Utilisé pour la détection des fraudes, la surveillance des systèmes, la détection de pannes, etc.

1.8 Conclusion

Ce chapitre a introduit les bases des Systèmes de Transport Ferroviaire Intelligents (STFI), en mettant en lumière les systèmes CBTC et TCCBTC. Nous avons exploré leurs architectures et les enjeux de sécurité associés. En outre, nous avons discuté de l'impact croissant de l'Intelligence Artificielle dans l'optimisation et la sécurisation des opérations ferroviaires. Le prochain chapitre approfondira spécifiquement les aspects de la sécurité dans les systèmes de transport ferroviaire.

2

ÉTUDE DE LA SÉCURITÉ DANS LES SYSTÈMES DE TRANSPORT FERROVIAIRE

2.1 Introduction

Le domaine des systèmes de transport ferroviaire intelligents représente un point d'intersection crucial entre l'innovation technologique et la nécessité impérieuse d'assurer la sécurité des passagers et des infrastructures. Dans ce chapitre, nous nous penchons sur l'état de l'art de la sécurité des systèmes de transport ferroviaire intelligents, explorant les progrès récents, les défis persistants et les solutions prometteuses dans ce domaine critique.

2.2 Synthèse des travaux connexes

Dans le domaine de la sécurité des système de transport férroviaire, il existe de nombreuses approches utilisées pour identifier les attaques.

Dans ce qui suit, nous étudions le principe de fonctionnement des travaux qui ont été menées afin de détecter les comportements anormaux.

2.2.1 Signal Jamming Attacks Against Communication-Based Train Control : Attack Impact and Countermeasure

Les auteurs de l'article [17] traitent les impacts des attaques de brouillage sur les systèmes de communication ferroviaire basés sur le contrôle des trains par communication (CBTC). Le problème réside dans la vulnérabilité de ces systèmes aux attaques de brouillage, en particulier dans le contexte de la communication guidée au milieu ouvert, où les répéteurs peuvent amplifier les signaux indésirables. Les chercheurs proposent une solution qui se concentre sur l'augmentation du temps de trajet des trains et du temps d'attente et de congestion des passagers. Ils ont utilisé une approche de cosimulation intégrant un modèle de mouvement des trains sous différents nœuds de signalisation (MBS et FBS) et du canal de communication sans fil sous les modes de communication en milieu ouvert et en milieu guidé. Les résultats ont mis en évidence des impacts sévères du brouillage, en particulier dans les systèmes de communication guidés en milieu ouvert, en expliquant l'amplification du signal par les répéteurs. Pour atténuer ces attaques, les chercheurs ont proposé une stratégie FHSS (Frequency Hopping Spread Spectrum) et ont évalué son efficacité à l'aide d'une plateforme de test basée sur SDR (Software Defined Radio). Les résultats ont montré une amélioration significative du SINR (Signal-to-Interface Noise Radio) avec l'implication de la méthode FHSS, réduisant efficacement l'impact des attaques de brouillage.

Discussion et critiques

La solution FHSS pour atténuer les attaques de brouillage dans les systèmes ferroviaires présente des défis techniques liés à la complexité de mise en œuvre et la synchronisation, à une consommation élevée de bande passante et à des coûts d'infrastructure. De plus, elle nécessite des tests en conditions réelles et pourrait être contournée par des attaquants sophistiqués.

2.2.2 A Novel Intrusion Detection Method in Train-Ground Communication System

L'article [18] présente une méthodologie visant à détecter les attaques par déni de service dans les systèmes de communication train-sol basés sur le contrôle des trains par communication (CBTC) ; les auteurs cherchent à identifier et à atténuer les effets en utilisant un algorithme d'apprentissage automatique appelé AdaBoost multi-classe. Cet algorithme est entraîné à reconnaître les schémas de comportement associés aux opérations normales et aux attaques. La classification multi-boost améliorée contient huit phases : La première phase consiste à initialiser chaque exemple dans l'ensemble des données de manière égale. La deuxième phase permet de sélectionner un classifieur faible pour minimiser l'erreur pondérée sur l'ensemble des données. Les poids des exemples sont mis à jour pour donner plus de poids aux exemples mal classés par le classifieur faible, afin de

les traiter de manière plus intensive lors de l'itération suivante. Les classifieurs faibles sont pondérés en fonction de leur performance lors de la classification pour chaque classe. Les classifieurs qui ont une performance supérieure contribuent davantage à la décision finale. La cinquième étape consiste à attribuer les exemples à la classe pour laquelle le classifieur faible associe la plus grande confiance, et l'erreur totale du modèle sera calculée en fonction de la performance combinée des classifieurs faibles pour chaque classe. Les étapes 2 à 6 sont répétées pour chaque classe jusqu'à l'obtention des meilleures performances. Une fois que toutes les itérations sont terminées, le modèle final est construit. Par Dans cet article, quatre indicateurs statistiques sont adoptés pour évaluer la performance de détection : Le taux de détection, le taux de faux positifs, le taux de faux négatifs et le taux d'erreur d'identification des attaques (AIER). L'algorithme de statistique de détection des anomalies de référence [3] (H. ALIPOUR, Y. B., AL-NASHIF, P...) et l'algorithme de classification multi-boost traditionnel sont utilisés comme expériences de comparaison. Par Les résultats de la simulation montrent que l'algorithme de classification multi-boost proposé présente de meilleures performances de détection pour les attaques DOS, car il présente des performances de détection élevées avec des taux de faux positifs et de faux négatifs généralement faibles pour différentes catégories d'attaques dans le contexte de la communication train-sol.

Discussion et critiques

L'article présente une méthode de détection d'intrusion utilisant AdaBoost multi-classe amélioré, mais la complexité algorithmique et la variabilité des performances en fonction des attaques spécifiques soulèvent des préoccupations. De plus, les taux de faux positifs et de faux négatifs, bien qu'améliorés, nécessitent une surveillance continue.

2.2.3 A Novel Intrusion Detection Model Using a Fusion of Network and Device States for Communication-Based Train Control Systems

Yajie Song, Bing Bu et Li Zhu [19] ont proposé une nouvelle méthode de détection d'intrusion qui prend en compte à la fois l'état des réseaux et celui des équipements pour identifier si l'anomalie est causée par des cyberattaques ou par des défaillances du système. Un modèle de Markov caché (HMM) est utilisé pour fusionner les informations provenant de différents modèles afin de prendre des décisions sur les résultats de la détection.

L'approche comprend plusieurs modèles de détection, dont un classificateur basé sur les modèles de Markov cachés (HMM). Cette méthode repose sur plusieurs étapes clés. Tout d'abord, une analyse approfondie des différents types de cyberattaques est réalisée, incluant l'examen des attaques potentielles telles que le déni de service (DoS) et les attaques par injection de données (DIA). Sur la base de cette analyse, le modèle de détection basé sur les états des dispositifs (DAD) est élaboré. Les résultats des modèles de détection des anomalies du réseau (NAD) et des dispositifs (DAD) sont ensuite fusionnés pour

former un système de détection d'intrusion complet. Un classificateur HMM est utilisé pour différencier les cyberattaques des défaillances aléatoires du système.

Le HMM est entraîné à partir des données collectées pour prédire les états cachés du système à partir des observations. À travers des expérimentations limitées, il a été démontré que l'IDS proposé peut atteindre un taux de détection de 97,64 % tout en maintenant un taux de fausses alertes inférieur à 2,66 %.

Discussion et critiques

L'article propose une méthode innovante pour la détection des intrusions dans les systèmes CBTC, combinant des modèles de détection basés sur l'état du réseau et des dispositifs. Cependant, il présente certaines limites, notamment une performance inférieure pour la détection des attaques de falsification de données comparée à d'autres types d'attaques. De plus, l'approche repose fortement sur des données simulées, ce qui peut ne pas refléter parfaitement les conditions réelles. Enfin, la solution ne mentionne pas le temps de réponse détaillé pour la détection et la réaction aux attaques, ce qui est crucial pour la sécurité des systèmes CBTC en temps réel.

2.2.4 Improve Safety and Security of Intelligent Railway Transportation System Based on Balise Using Machine Learning Algorithm and Fuzzy System

Les auteurs de l'article [20] ont pour objectif d'étudier les cyber-attaques sur les balises pouvant entraîner des troubles physiques. Ils proposent des méthodes de détection des attaques et de pannes pour remédier aux limitations et contraintes des infrastructures ferroviaires. Un système de contre-mesure flou est développé pour améliorer la sécurité des trains contre les cyber-attaques, qu'elles soient connues ou inconnues. Dans cette approche, les données sensorielles du train servent de caractéristiques pour les algorithmes d'apprentissage automatique tels que ANFIS, SVM et le perceptron multicouche (MLP) afin de classifier les opérations standard et les attaques. Ce système est placé dans une couche intermédiaire (ordinateur virtuel) où il analyse les données reçues de l'ordinateur vital, surveille les différents capteurs et enregistre ses observations dans un journal d'audit. Ces enregistrements sont ensuite analysés pour détecter les opérations anormales. Après la détection d'attaque, un signal d'alerte est déclenché et la réaction est effectuée en fonction des signaux et des conditions du train par un contrôleur flou. L'algorithme ANFIS a montré une précision de détection élevée, atteignant 92% d'exactitude et 91% de précision. Cette approche proposée offre une méthode efficace pour garantir l'intégrité et la fiabilité des systèmes de contrôle des trains basés sur les balises.

Discussion et critiques

L'approche utilise des techniques avancées d'intelligence artificielle. Cependant, cer-

taines lacunes méritent d'être notées. Tout d'abord, bien que l'utilisation de méthodes d'apprentissage automatique telles que l'ANFIS et le MLP soit pertinente pour la détection d'anomalies, l'article ne fournit pas suffisamment de détails sur la performance de ces algorithmes dans des scénarios réels de déploiement. De plus, la généralisabilité de l'approche proposée à d'autres environnements ferroviaires ou à des types d'attaques différents n'est pas clairement discutée. Une évaluation plus approfondie de la robustesse et de la fiabilité de l'algorithme de détection des attaques est nécessaire pour garantir son efficacité dans des situations opérationnelles réelles.

2.2.5 Data Integrity Threats and Countermeasures in Railway Spot Transmission Systems

Dans l'article [21], les auteurs discutent les menaces de l'intégrité des données pour les modules de transmissions de balises. Pour atténuer ce risque, ils proposent une solution pratique à deux niveaux : Une situation cryptographique pour protéger l'intégrité des informations au niveau du dispositif et un contrôleur de vitesse de train hybride au niveau du système. La première contre-mesure consiste à une méthode légère et peu coûteuse d'authentification qui permet aux trains en passage de détecter toute altération des données des télégrammes fournis par les balises eurobalises. La seconde contre-mesure est un contrôleur de vitesse résilient conçu pour les applications de contrôle d'arrêt de train. Ce contrôleur exploite un schéma d'authentification et ajoute une couche de sécurité supplémentaire en fournissant une stratégie de réponse opérationnellement viable en cas de données incorrectes. Ces deux contre-mesures sont conçues pour être applicables aux systèmes existants, sans nécessiter de matériel supplémentaire ou de capteurs. La simulation démontre que les deux stations proposées sont efficaces pour sécuriser les systèmes de transmission ferroviaire contre les attaques, mais elles nécessitent des stratégies de réponse appropriées pour atténuer les conséquences des attaques sur les opérations ferroviaires.

Discussion et critiques

L'approche est considérée comme forte dans la sécurisation des systèmes de transmission de point de balise ferroviaire. Cependant, elle présente plusieurs points critiques. Tout d'abord, bien que les contre-mesures soient conçues pour être légères et intégrables aux systèmes existants, elles pourraient nécessiter des ajustements pour s'adapter à différentes infrastructures ferroviaires. De plus, la détection des attaques et la réaction en temps réel peuvent être affectées par des variations de conditions environnementales ou de performances du système, ce qui nécessiterait une analyse approfondie de la robustesse de l'approche dans différentes situations. Enfin, la mise en œuvre pratique des contre-mesures et leur déploiement à grande échelle pourraient nécessiter une coordination et une collaboration étroites entre les différents acteurs du secteur ferroviaire.

2.2.6 Security Architecture for Secure Train Control and Monitoring System

L'article [22] présente l'architecture Securebox, une solution visant à renforcer la sécurité du système de contrôle de train (TCMS). Cette architecture repose sur l'utilisation de modules matériels de sécurité (HSM) et de protocoles de chiffrement avancés pour sécuriser les communications et les données dans les systèmes de TCMS.

La Securebox se compose de quatre fonctions principales : la gestion du réseau, la gestion du buffer, la gestion des données et la gestion de la sécurité. Ensemble, ces fonctions garantissent la confidentialité, l'intégrité et l'authenticité des données échangées entre les trains et les serveurs centraux.

L'évaluation de la performance et de la sécurité de l'architecture Securebox, comprend des tests en boîte noire et en boîte blanche. Les résultats des tests montrent que l'utilisation des modules HSM n'a pas affecté les performances du TCMS, avec des temps de transmission et de chiffrement négligeables. De plus, l'utilisation de l'algorithme AES a démontré une amélioration significative des performances par rapport à 3DES, tout en assurant une protection contre les attaques MITM et autres menaces potentielles.

Discussion et critiques

Bien que l'architecture SecureBox présente de nombreux avantages en matière de sécurité des communications dans les systèmes de gestion des trains, elle n'est pas exempte de limitations. La dépendance aux réseaux sans fil expose le système à des interruptions potentielles de transmission et à des attaques de déni de service, ce qui peut compromettre la disponibilité des données essentielles. En outre, la complexité de mise en œuvre et de gestion des clés de chiffrement peut nécessiter des ressources considérables et une expertise spécialisée. Enfin, malgré les mesures de sécurité robustes, le système reste vulnérable aux nouvelles menaces et doit être régulièrement mis à jour pour maintenir son efficacité face aux évolutions des cyberattaques.

2.2.7 Improving the Security of LTE-R for High-Speed Railway : From the Access Authentication View

L'article [23] présente un schéma d'authentification novateur pour les réseaux ferroviaires LTE-R, visant à sécuriser les communications entre les trains et les stations de base. Cette solution repose sur plusieurs éléments clés. Tout d'abord, elle utilise l'algorithme ECC-CLPS (Elliptic Curve Cryptography-Based Certificateless Proxy Signature) pour l'authentification, simplifiant les échanges de clés tout en garantissant une sécurité renforcée. Ensuite, un puzzle basé sur le hash est intégré pour contrer les attaques par déni de service (DoS), offrant une protection robuste contre les tentatives de surcharge

du système. De plus, une technique de pré-génération de clés est employée pour optimiser le processus d'authentification, réduisant les délais et les coûts associés. L'analyse de sécurité exhaustive démontre la résilience du schéma face à diverses attaques potentielles, tandis que l'évaluation de performance met en évidence son efficacité en termes de coûts de calcul et de communication. Ce schéma d'authentification représente une avancée significative dans la sécurisation des communications dans les réseaux ferroviaires, offrant une solution robuste et prometteuse pour répondre aux exigences de sécurité spécifiques de ces environnements sensibles.

Discussion et critiques

L'approche proposée, bien qu'innovante et sécurisée grâce à l'utilisation de l'algorithme ECC-CLPS et des puzzles de hachage, présente une complexité computationnelle et une consommation d'énergie accrues, posant des défis pour les dispositifs à ressources limitées. De plus, l'absence de validation pratique en environnement réel laisse planer des incertitudes quant à son efficacité opérationnelle.

Le tableau 2.1 résume les approches étudiées.

Article	Problématique	solution	approche
[1]	La vulnérabilité des CBTC aux attaques de brouillage	l'application du FHSS sur la paire émetteur-récepteur, également sur les répéteurs sans fil.	l'utilisation de la technique de saut de fréquence en séquence directe (solution technique)
[2]	La sécurisation des communications, détecter et contrer les attaques.	Proposition d'une méthode de détection des intrusions.	Utilisation de l'algorithme AdaBoost.
[3]	La nécessité de protéger les systèmes de contrôle de la circulation et de la signalisation basés sur CBTC des cyberattaques.	Un système de détection d'intrusion	Une combinaison de modèles de détection d'intrusion fondés sur l'état du réseau et des appareils, ainsi qu'un classificateur (HMM).
[4]	La détection et la prévention des attaques potentielles visant les systèmes de contrôle ferroviaire basés sur les balises.	Méthodes de détection des attaques et des pannes	Combiner des techniques d'apprentissage automatique avec des données, les trains et des balises.
[5]	la sécurisation des systèmes de transmission de balises ferroviaires contre les attaques de manipulation et de clonage.	Méthodes d'authentification légères et un contrôleur de vitesse résilient.	Incorporation d'une étiquette d'authentification dans les télégrammes transmis par les balises et développement d'un contrôleur de vitesse résilient.
[6]	La nécessité de renforcer la sécurité des TCMC.	Implémentation de l'architecture SecureBox.	Intègre des modules matériels sécurisés (SecureBox HSM) dans les TCMS.
[7]	Sécurisation des communications dans le contexte des réseaux LTE-R ; concevoir une solution d'authentification.	Un schéma d'authentification d'accès basé sur l'algorithme ECC-CLPS.	combinaison de plusieurs techniques de sécurité et de cryptographie.

TABLE 2.1 – Tableau récapitulatif.

2.3 Classification des travaux

Dans cette section, nous classifions les travaux analysés en utilisant deux critères. le premier représente les types de communications visés (communication entre les trains, entre les trains et la station de base et entre les trains et les balises), et le deuxième critère représente les approches utilisées. La table 2.2 montre notre classification.

les approches	communication train-balise	communication train-station de base
la technique de saut de fréquence (FHSS)	/	[1]
Apprentissage automatique	[4] [3]	[2] [4]
Modèle de Markov	[3]	/
Méthodes formelles	[5]	[7]
Modules matériels sécurisés et des mécanismes de cryptage	/	[6]
Courbes Elliptique	/	[7]

TABLE 2.2 – Classification

2.4 Comparaison et synthèse

le tableau 2.3 illustre une comparaison entre les approches de détection de comportement anormal selon les résultats obtenue et son efficacité.

Approche	Avantages	Inconvénients	Efficacité/Résultats
Algorithme AdaBoost	Haute précision de détection	Variabilité des performances selon les types d'attaques	Performances de détection élevées pour les attaques DOS avec faibles taux de faux positifs et faux négatifs
Modèle de Markov caché (HMM)	Combine les états des réseaux et des dispositifs	Dépendance élevée aux données simulées	Taux de détection de 97,64% et taux de fausses alertes inférieur à 2,66%
Apprentissage automatique (ANFIS, SVM, MLP)	Haute précision dans la détection d'anomalies	Manque d'évaluation en conditions réelles	Précision de détection atteignant 92% pour ANFIS
Méthodes d'authentification et contrôleurs de vitesse	Efficacité dans la protection contre les altérations de données	Nécessitent des ajustements pour différentes infrastructures	Solution légère et efficace pour sécuriser les systèmes de transmission
Architecture SecureBox	Protection robuste contre les cyberattaques	Implémentation nécessitant une expertise et des ressources significatives	Amélioration significative des performances et protection contre les attaques MITM
Schéma d'authentification basé sur ECC-CLPS	Sécurité renforcée et simplification des échanges de clés	Complexité computationnelle et consommation d'énergie accrues	Résilience démontrée contre diverses attaques et efficacité en termes de coûts de calcul et de communication
Technique FHSS	Améliore la résilience aux attaques de brouillage	Nécessite une infrastructure complexe et une synchronisation précise	Réduction significative de l'impact des attaques de brouillage

TABLE 2.3 – Comparaison des travaux analysés.

Les approches de sécurité des systèmes de transport ferroviaire intelligents varient en termes d'efficacité et de contraintes. La technique FHSS réduit les attaques de brouillage mais demande une infrastructure complexe. AdaBoost détecte efficacement les attaques DOS avec des performances variables selon le type d'attaque. Les modèles de Markov cachés offrent une haute détection mais dépendent des données simulées. Les méthodes d'apprentissage automatique et les contre-mesures légères montrent une précision élevée

et une intégration facile, respectivement. L'architecture SecureBox et le schéma ECC-CLPS renforcent la sécurité, bien qu'ils nécessitent des ressources et soient énergivores. Chaque solution présente des avantages spécifiques adaptés à des contextes opérationnels particuliers.

2.5 Conclusion

En somme, ce chapitre a exploré l'état de l'art en matière de sécurité dans les systèmes de transport ferroviaire intelligents. Nous avons analysé diverses approches pour atténuer les menaces potentielles, allant des attaques de brouillage à la détection d'intrusions et à la sécurisation des communications entre les trains et les infrastructures de contrôle. Chaque solution, qu'elle repose sur des algorithmes de machine learning, des techniques de cryptographie ou des architectures matérielles sécurisées, présente des avantages spécifiques, mais également des limitations. Les connaissances acquises servent de fondement pour le développement de stratégies plus robustes et adaptées à la sécurité des systèmes de transport ferroviaire intelligents. Dans le chapitre suivant, nous proposerons une solution pour renforcer l'authentification et l'intégrité des informations échangées entre les trains intelligents, en réponse aux défis identifiés.

3

A Hybrid Approach for secure communications in TC-CBTC system

3.1 Introduction

Dans le chapitre précédent, nous avons examiné des approches de sécurité des systèmes ferroviaires, en soulignant les problèmes principaux. Nous avons identifié les aspects critiques tels que l'authentification et l'intégrité des données échangées, qui sont insuffisamment traités dans les travaux proposés dans la littérature.

Dans le cadre de notre travail, nous avons choisi de concentrer nos efforts sur le développement d'une technique d'authentification et d'intégrité des données pour le système TC-CBTC (pour Train-Centric Communication-based Train Control), compte tenu de son importance cruciale dans la sûreté et la fiabilité des transports ferroviaires. Dans ce chapitre, nous présentrons notre approche de sécurité pour le système TC-CBTC.

3.2 Motivations

L'Autorité de Mouvement (MA) dans le système TC-CBTC (Train Control - Communications Based Train Control) vise à améliorer la sûreté et l'efficacité des opérations ferroviaires. L'autorité de mouvement est essentielle pour garantir que chaque train opère dans les limites de sécurité définies et respecte les règles de circulation établies. En incorporant la MA au sein du système TC-CBTC, nous permettons aux trains de recevoir des instructions précises et fiables concernant leur vitesse, leur trajectoire et leur espacement par rapport aux autres trains. Cela est essentiel pour éviter les collisions et garantir un flux de trafic ferroviaire fluide et sécurisé.

Cependant, l'introduction de l'autorité de mouvement dans les systèmes TC-CBTC n'est pas sans risques. Les données échangées pour calculer la MA doivent être sécurisées contre les cyberattaques, les intrusions malveillantes et les manipulations de données. Les risques potentiels incluent la falsification des informations de position, la création de fausses autorisations de mouvement et la perturbation des communications entre les trains. Par conséquent, il est impératif de mettre en œuvre des mécanismes de sécurité robustes pour garantir l'intégrité et l'authenticité des données échangées.

La motivation derrière notre proposition de solution réside dans la nécessité de garantir la sécurité et la fiabilité des systèmes de transport ferroviaire dans un environnement de plus en plus complexe et interconnecté. En assurant une authentification robuste pour les communications entre les trains et le calcul d'une MA, nous pouvons prévenir les risques potentiels tels que les cyberattaques, les intrusions malveillantes ou les manipulations de données, qui pourraient compromettre la sécurité et l'efficacité des opérations ferroviaires. De plus, en garantissant l'intégrité des informations échangées, notre solution peut réduire les risques d'incidents et d'accidents ferroviaires, assurant ainsi la sécurité des passagers et du personnel ferroviaire. En renforçant la confiance dans les systèmes de transport ferroviaire, notre solution favorise l'adoption continue de ces technologies innovantes et contribue à la durabilité à long terme des infrastructures de transport.

3.3 Modèle d'attaque

la figure 3.1 représante le modèle d'attaque servant de base à notre proposition ; ou l'attaquant intercepte et potentiellement altère les communication entre les deux trains ; le train T_i tente d'nvoyer un message au train T_{i-1} mais l'envoi échoue, l'attaquant remplace le message M par un autre message illégitime M'.

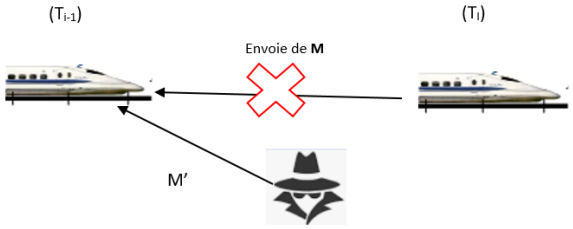

FIGURE 3.1 – Notre modèle d'attaque.

3.4 Notre modèle de communication sécurisée

Notre modèle de communication sécurisée est composé de trois phases distinctes, la phase d'authentification pour garantir l'identité des trains, la phase de vérification de l'intégrité des données pour s'assurer que les informations échangées n'ont pas été altérées, et la phase de prise de décision pour déterminer les actions à entreprendre.

3.4.1 Description

la Figure 3.2 résume la solution proposée. Le schéma récapitulatif présente le processus d'authentification et de vérification de l'intégrité des données entre trains. Il débute par la demande de connexion (D), où un train émetteur (T_i) envoie une demande signée numériquement au train récepteur. Ce dernier vérifie l'authenticité de la demande en déchiffrant la signature et en validant le certificat numérique du train émetteur. Une fois toutes les vérifications effectuées, une réponse d'acceptation est envoyée, confirmant la connexion sécurisée. En parallèle, le module de vérification de l'intégrité des données, basé sur des techniques d'apprentissage automatique, entre en action. Il extrait les caractéristiques du message reçu et les compare aux schémas de communication typiques. En cas de détection d'anomalies significatives, des alertes sont générées pour signaler des attaques potentielles. Dans le cas contraire, le système calcule l'autorité de mauvement. Les notations utilisées sont données dans le tableau 3.1.

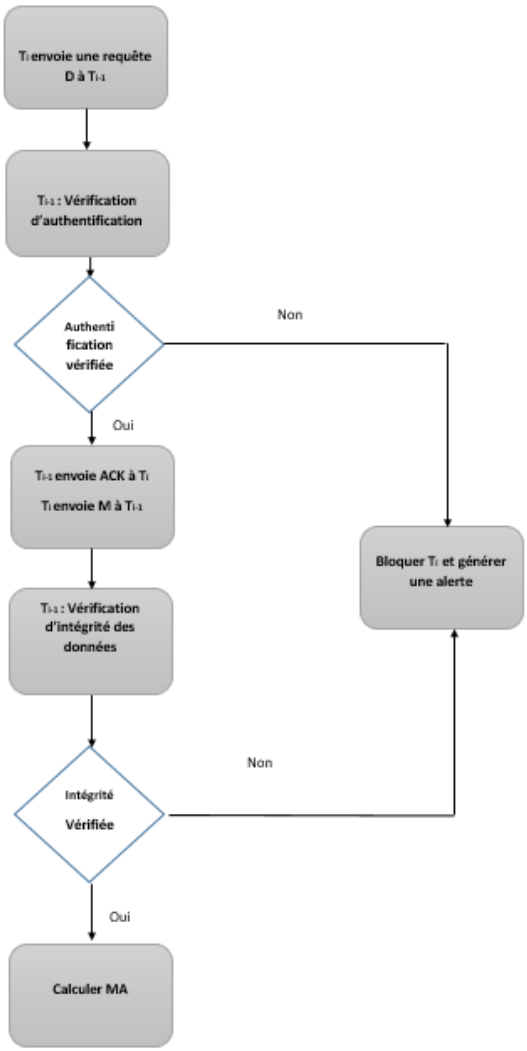

FIGURE 3.2 – Schéma récapitulatif de notre proposition.

Paramètre	Signification
T_i	le train émetteur
T_{i-1}	le train récepteur
D	la demande de connexion
ACK	accusée de réception de D et acceptation
M	l'ensemble de données envoyé
V	vitesse du train
T	le temps
(x,y)	la position horizontale du train T_i

TABLE 3.1 – Notations utilisées.

3.4.2 Module d'authentification

Pour sécuriser les communications entre les trains dans le système TC-CBTC, il est crucial d'établir une clé partagée qui servira à chiffrer et déchiffrer les signature numérique. Nous utilisons le protocole Diffie-Hellman pour réaliser un échange de clés sécurisé entre deux trains. Les deux trains conviennent d'un grand nombre premier p et d'une base g, qui peuvent être partagés publiquement et définis par le système de communication ferroviaire. Le train T_i choisit un nombre secret aléatoire a et calcule sa clé publique $A = g^a \mod p$, qu'il envoie à T_{i-1}. Simultanément, T_{i-1} choisit un nombre secret aléatoire b et calcule sa clé publique $B = g^b \mod p$, qu'il envoie à T_i. Le train T_i reçoit la clé publique B de T_{i-1} et calcule la clé partagée $K = B^a \mod p$, tandis que T_{i-1} reçoit la clé publique A de T_i et calcule la clé partagée $K = A^b \mod p$.

Chaque train signe numériquement sa clé publique avant de l'envoyer pour garantir son authenticité. T_i signe A avec sa clé privée et l'envoie à T_{i-1}, tandis que T_{i-1} signe B avec sa clé privée et l'envoie à T_i. Chaque train vérifie la signature reçue avec la clé publique de l'autre train pour s'assurer de l'authenticité de la clé publique reçue.

Après cet échange de clés sécurisé, nous passons à l'authentification entre trains.

Lors de l'authentification entre trains, chaque message transmis comprend des données essentielles telles que la vitesse, le temps et les coordonnées horizontales (x, y),... du train émetteur. La formule mathématique 3.1. est utilisée pour illustrer la demande de connexion D.

$$D = (ID_i, Protocole)_S \tag{3.1}$$

La figure 3.3 représente le processus de demande de connexion. Lorsqu'un train T_i. souhaite établir une connexion sécurisée avec son précédent T_{i-1}, il lui envoie une demande de connexion D. Avant d'envoyer D, le T_i signe numériquement le message à l'aide de sa clé privée, assurant ainsi son authenticité. À la réception, le T_{i-1} vérifie la légitimité de la demande en déchiffrant la signature numérique à l'aide de la clé publique du T_i. En parallèle, le certificat numérique du T_i est validé en utilisant la clé publique de l'autorité de certification. Si toutes ces vérifications sont concluantes et que le message n'a pas été altéré en transit, le T_{i-1} envoie une réponse d'acceptation, confirmant ainsi l'établissement d'une connexion sécurisée. Ce processus d'authentification, basé sur des techniques cryptographiques, garantit que seuls les trains autorisés peuvent établir des connexions sécurisées, renforçant ainsi la sécurité des échanges de données entre les trains.

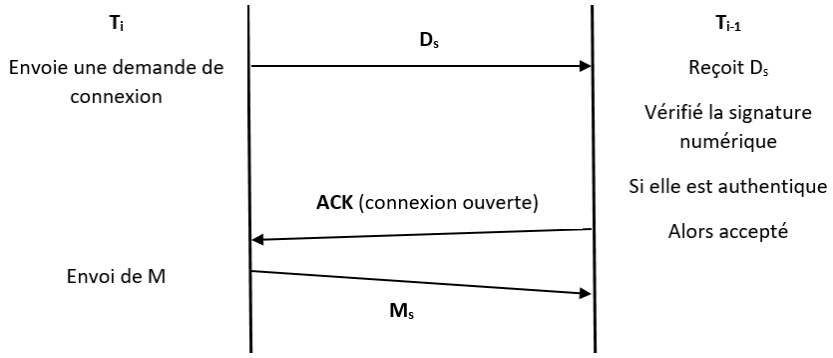

FIGURE 3.3 – Demande de connexion.

3.4.3 Module de vérification de l'intégrité de données

Le module de vérification de l'intégrité des données développé, repose sur des techniques d'apprentissage automatiques (ML pour Machine Learning). Le module se compose de trois étapes successives pour maximiser les performances de détection et de prise de décision (voir la figure 3.4). Un processus d'apprentissage hors ligne du module est effectuée.

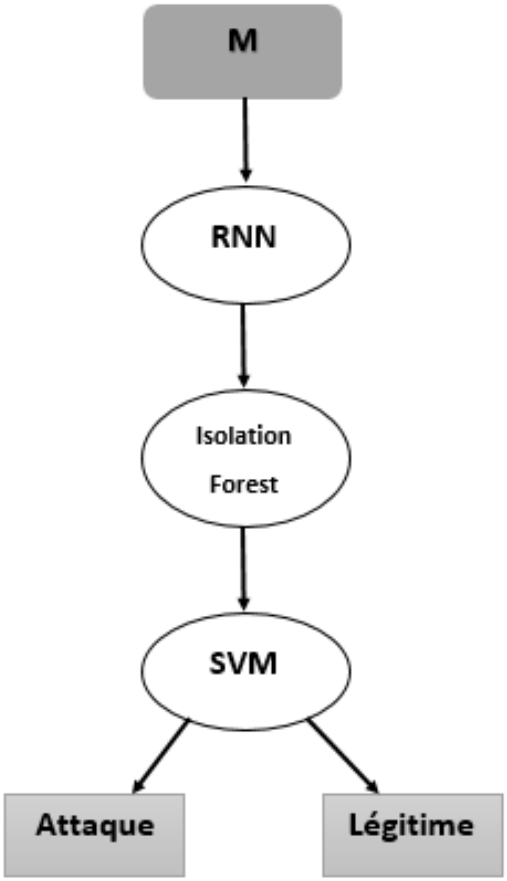

FIGURE 3.4 – Vérification de l'intégrité des données échangées.

Lorsqu'un message $M = (V, T, x, y, nb\_a, gare\_d, gare\_a)$ est envoyé, comprenant des informations telles que la vitesse (V), le temps (T), les données de position horizontale (x, y) du T_i, le nombre d'arrêts (nb_a), la gare de départ (gare_d) et la gare d'arrivée (gare_a), ce module entre en action du côté du T_{i-1}. En recevant le message, le modèle extrait les caractéristiques pertinentes du message afin de les comparer aux schémas de communication typiques appris précédemment.

Tout d'abord, un réseau neuronal récurrent (RNN pour Recurrent Neural Network) est utilisé pour modéliser les dépendances temporelles dans les données de base des trains, telles que la vitesse, la distance, le temps et les positions (x, y). L'entrée du RNN est l'ensemble du dataset comprenant les séquences de données temporelles des trains, et sa sortie est une représentation codée de ces séquences, capturant les relations temporelles entre les différents points de données.

Ensuite, les sorties du RNN sont traitées par l'algorithme Isolation Forest. L'entrée de l'Isolation Forest est cette représentation codée fournie par le RNN, et sa sortie est une évaluation de chaque point de données quant à sa conformité aux schémas normaux,

identifiant ainsi les schémas inhabituels ou anormaux dans le comportement du train.

Enfin, les données affinées par Isolation Forest sont classées par l'algorithme machine à vecteur de support (SVM). L'entrée du SVM est les résultats de l'Isolation Forest, et sa sortie est une classification précise de chaque point de données en termes de normalité ou d'anormalité du comportement du train.

Si le modèle détecte des déviations significatives par rapport aux schémas typiques, il peut alors alerter sur d'éventuelles anomalies ou attaques. En revanche, si le message est conforme aux schémas de comportement normal, il est considéré comme authentique et peut être traité en conséquence.

3.4.4 Prise de décision

Une fois les phases d'authentification de l'émetteur et de vérification de l'intégrité des données achevées, le modèle proposé doit prendre une décision concernant le message reçu. Dans ce qui suit, nous décrivons en détail comment la phase de prise de décision peut se dérouler.

3.4.4.1 Cas d'attaque

Lorsqu'une attaque est détectée dans la communication, une série d'actions est déclenchée pour gérer la situation de manière efficace. Tout d'abord, le système va transmettre les informations relatives à l'attaque à la station de base responsable de la gestion du système ferroviaire. Ces informations comprennent des détails sur la nature de l'attaque, les données associées au message compromis.

Une fois ces informations reçues, la station de base diffuse l'identité et les coordonnées du train compromis à tous les autres trains du réseau pour l'isoler du reste du réseau.

En outre, un processus de mise à jour de l'apprentissage du module de vérification de l'intégrité sera initie sur place sur des données d'attaque actualisées. Pendant ce processus, les paramètres du modèle sont ajustés et optimisés pour mieux s'adapter aux caractéristiques des nouvelles données et améliorer la capacité de détection des attaques futures.

3.4.4.2 Cas normal

Dans le cas d'une communication normale sans aucune attaque détectée, le train calcule la distance entre le train émetteur et le train récepteur (MA) à l'aide des données

de status fournies dans le message. En utilisant ces coordonnées, le système applique des formules mathématiques appropriées pour déterminer la distance entre les deux trains.

3.5 Conclusion

Dans ce chapitre, nous avons introduit notre approche HASC (Hybrid Approach for Secure Communications in TC-CBTC system), qui repose sur l'authentification et la vérification de l'intégrité des données pour sécuriser les communications entre les trains. En garantissant la fiabilité des échanges, cette solution vise à prévenir les incidents et à protéger la sécurité des passagers et du personnel ferroviaire, contribuant ainsi à une plus grande sûreté et fiabilité des transports ferroviaires dans un environnement complexe et interconnecté. Le chapitre suivant abordera la simulation et l'évaluation des performances de notre modèle.

4

Evaluation de performances

4.1 Introduction

Ce chapitre est dédié à la simulation et à l'analyse des performances de notre modèle de communication sécurisée dans le cadre du système TC-CBTC. Notre évaluation vise à mesurer la précision, le rappel, l'exactitude et le F1-score de notre approche. Nous commençons par présenter le dataset utilisé ainsi que les étapes réalisées lors du prétraitement des données pour assurer leur qualité et leur pertinence. Ensuite, nous décrivons la création du modèle, en mettant en lumière les techniques d'apprentissage automatique et de cryptographie intégrées pour garantir la sécurité des communications. Nous expliquons également les métriques de performance utilisées pour évaluer notre solution. Enfin, nous fournissons une interprétation détaillée des résultats obtenus à partir de la simulation de notre modèle.

4.2 Mise en oeuvre

4.2.1 Environnement de développement

Un environnement de développement est un ensemble d'outils logiciels qui facilite le développement de logiciels en offrant des services complets pour les programmeurs. Nous utilisons Visual Studio Code (VS Code) pour le développement de notre solution ;

qui est un éditeur de code source développé par Microsoft, disponible gratuitement et en open source. Il est léger mais puissant et offre des fonctionnalités robustes pour le développement de logiciels [24].

4.2.2 Langage de programmation

Le langage de programmation utilisé dans le cadre de notre simulation est le langage Python, qui est un langage interprété, de haut niveau et à usage général. La philosophie de conception de Python met l'accent sur la lisibilité du code grâce à l'utilisation notable de l'indentation significative. Ses constructions linguistiques et son approche orientée objet visent à aider les programmeurs à écrire du code clair et logique pour des projets de petite et grande envergure. Python est typé dynamiquement et dispose d'un ramasse-miettes pour la gestion automatique de la mémoire. Il supporte plusieurs paradigmes de programmation, y compris la programmation structurée (notamment procédurale), orientée objet et fonctionnelle. Python est souvent décrit comme un langage "batteries incluses" en raison de sa bibliothèque standard complète [25].

4.2.3 Bibliothèques

Les bibliothèques utilisées sont :

- **sklearn (scikit-learn) :** pour Prétraitement des données, création et évaluation de notre modèles de machine learning.
- **tensorflow.keras :** pour Construction de RNN, entraînement et évaluation de modèleS.
- **seaborn :** pour la visualisation des données pour l'analyse exploratoire et la présentation des résultats.
- **pandas :** pour la manipulation des données en vue de l'entraînement et l'évaluation de notre modèle.
- **datetime :** pour la manipulation de dates et d'heures, calculs et formats de temps.
- **joblib :** pour sauvegarde et chargement de modèles, exécution de tâches parallèles.
- **cryptography.hazmat.primitives :** pour l'implémentation de fonctions cryptographiques ; la signature numérique, les chiffrement/déchiffrement, et la gestion des clés.

4.2.4 Schéma récapilatif

la figure 4.1 représente les étapes de la phase d'apprentissage ;

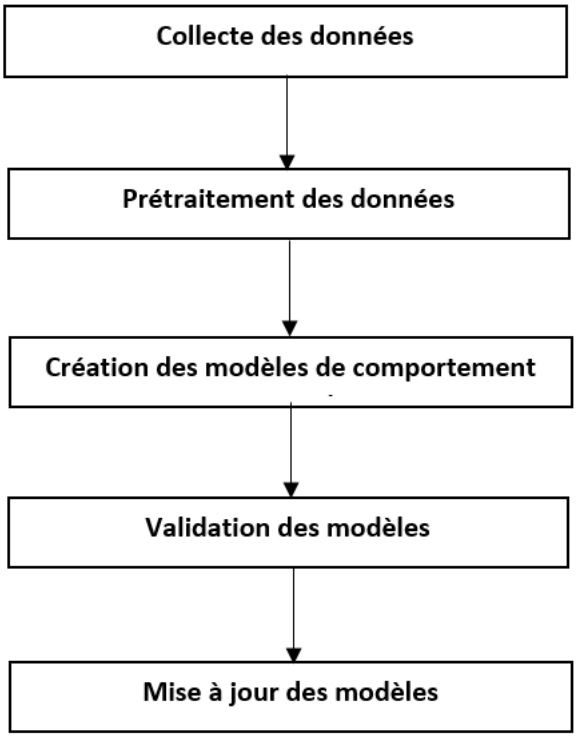

FIGURE 4.1 – Etapes d'apprentissage.

4.2.5 Collecte de données

Les ensembles de données historiques sont collectés à partir des échanges de messages entre les trains dans des environnements de test contrôlés. Ces données servent à construire des modèles de comportement normal pour chaque train et à identifier les schémas de communication typiques. Après la collecte, nous effectuons un prétraitement des données pour éliminer les valeurs aberrantes, gérer les données manquantes et sélectionner les caractéristiques pertinentes.

4.2.5.1 Construction du dataset utilisé

Dans un premier temps, nous avons rassemblé deux ensembles de données : l'un provenant d'un site Web open source (www.ressources.data.sncf.com, consulté le 10 mai 2024) intitulé "Vitesse maximale nominale sur ligne" et l'autre provenant d'un autre site web open source (static.data.gouv.fr, consulté le 10 mai 2024) intitulé "Horaires des lignes TER". Nous avons complété les données manquantes à l'aide de l'application de réservations de chaque ligne (horaires-de-trains.fr, consultée le 11 mai 2024). L'ensemble des données représente le contenu des messages échangés entre les trains.
Le tableau 4.1 montre les caractéristiques de l'ensemble de données.

Attributs	Description	Type
id_train	identifiant du train	charactère (input)
V_MAX	la vitesse maximale	reel (input)
Geo Point	la géolocalisation du train	reel (input)
LABEL	la classe des données	booléen (output)

TABLE 4.1 – Descreption du dataset.

Les autres attributs représente les informations concernat la gare de depart et la gare d'arrivée ainsi que les heures de departs et celle d'arrivées.

4.2.6 Prétraitement des données

Le prétraitement des données de communication est une étape cruciale dans le traitement de données liées à la sécurité, qui vise à préparer les données brutes pour l'analyse et la modélisation. Cette étape comprend plusieurs étapes essentielles, telles que le nettoyage des données, la normalisation des données, la transformation des données et l'extraction de fonctionnalités.

Voici les étapes détaillées du prétraitement des données de message communiquer entre les trains :

1. **Nettoyage des données :** Cette étape vise à corriger les erreurs, supprimer les données aberrantes, remplacer les valeurs manquantes et éliminer les bruits. Les erreurs peuvent provenir de capteurs défectueux ou de perturbations environnementales, tandis que les données aberrantes peuvent être causées par des événements imprévus comme des attaques. Le remplacement des valeurs manquantes est crucial pour éviter les erreurs dans les analyses, et l'élimination des bruits réduit la variabilité des données. Les techniques couramment utilisées incluent la détection de données aberrantes, l'imputation de données manquantes et le filtrage des données.

   Les figure 4.2 et 4.3 représente les fonctions utilisées pour le nettoyage des données.

```
# Remplir ou supprimer les valeurs manquantes
data.fillna({
    'heure_depart_prevue': '00:00:00',
    'nom_gare_depart': 'inconnu',
    'id_gare_depart': -1,
    'nom_gare_arrivee': 'inconnu',
    'id_gare_arrivee': -1,
    'nombre_arrets': 0,
    'retard_max_en_secondes': 0,
    'retard_en_minutes_par_gare': 0,
    'heure_d\'arrivée': '00:00:00',
    'V_MAX': 120,
    'Geo Point à t0': 'inconnu'
}, inplace=True)

# Remplacer les valeurs problématiques dans la colonne 'retard_en_minutes_par_gare' par NaN
data['retard_en_minutes_par_gare'] = pd.to_numeric(data['retard_en_minutes_par_gare'], errors='coerce')

# Conversion des colonnes de temps en datetime
data['heure_depart_prevue'] = pd.to_datetime(data['heure_depart_prevue'], format='%Hh%M%S').dt.time
data['heure_d\'arrivée'] = pd.to_datetime(data['heure_d\'arrivée'], format='%Hh%M%S').dt.time

# Conversion des colonnes numériques
data['nombre_arrets'] = data['nombre_arrets'].astype(int)
data['retard_max_en_secondes'] = data['retard_max_en_secondes'].astype(int)
```

FIGURE 4.2 – Remplir ou supprimer les valeurs manqauntes.

```
# Supprimer les lignes avec des valeurs manquantes
data.dropna(subset=['geo_x', 'geo_y'], inplace=True)

# Supprimer les colonnes non pertinentes
data = data[['heure_depart_prevue', 'nombre_arrets', 'heure_arrivee', 'V_MAX', 'geo_x', 'geo_y', 'retard

# Supprimer les lignes avec des valeurs manquantes
data.dropna(inplace=True)
```

FIGURE 4.3 – Suppression.

2. **Normalisation des données :** Cette étape consiste à transformer les données pour qu'elles aient une échelle commune, ce qui est essentiel pour améliorer la performance des algorithmes de machine learning. La normalisation permet de réduire les biais dus aux différentes échelles de mesure et assure que chaque variable contribue de manière équitable à l'analyse. Les techniques courantes de normalisation incluent la mise à l'échelle des données entre 0 et 1, la standardisation (centrer et réduire les données) et la transformation logarithmique pour traiter les distributions asymétriques.

   La figure 4.4 représente la normalisation des données.

```
# Appliquer la normalisation
# Définir les colonnes à normaliser
cols_to_normalize = ['retard_max_en_secondes', 'retard_en_minutes_par_gare', 'V_MAX']

# Z-score normalization
scaler_standard = StandardScaler()
train_data_standard = train_data.copy()
test_data_standard = test_data.copy()

train_data_standard[cols_to_normalize] = scaler_standard.fit_transform(train_data[cols_to_normalize])
test_data_standard[cols_to_normalize] = scaler_standard.transform(test_data[cols_to_normalize])

# Min-Max normalization
scaler_minmax = MinMaxScaler()
train_data_minmax = train_data.copy()
test_data_minmax = test_data.copy()

train_data_minmax[cols_to_normalize] = scaler_minmax.fit_transform(train_data[cols_to_normalize])
test_data_minmax[cols_to_normalize] = scaler_minmax.transform(test_data[cols_to_normalize])
```

FIGURE 4.4 – Normalisation des données.

3. **Transformation des données :** Cette étape implique la modification des données brutes en un format plus approprié pour l'analyse et la modélisation. Elle comprend des techniques comme la création de nouvelles variables à partir des données existantes. La transformation des données améliore la qualité des modèles en rendant les relations entre les variables plus linéaires et en réduisant l'impact des outliers et des distributions asymétriques.

la figure 4.5 représente la transformation des données.

FIGURE 4.5 – Transformation des données.

4. **Extraction de fonctionnalités :** Cette étape consiste à identifier les caractéristiques importantes des données et à les extraire pour faciliter leur utilisation dans les modèles de conduite. Dans cette étape, nous avons extrait les coordonnées de géolocalisation, x et y, pour les intégrer dans notre modèle. Ces données sont essentielles pour capturer les mouvements et les positions des trains. En analysant les trajectoires et les comportements de conduite, nous pouvons mieux comprendre les contextes spatiaux des incidents et des performances. L'extraction de ces fonctionnalités géographiques fournit des informations précieuses pour améliorer la sécurité et l'efficacité des systèmes de transport ferroviaire intelligents.

La figure 4.6 représentre l'extraction des coordonnées de géolocalisation.

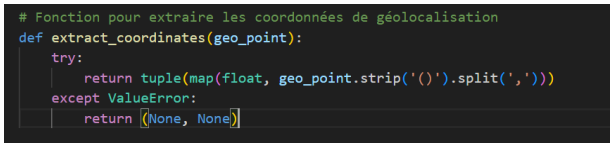

FIGURE 4.6 – Extraction des données.

4.3 Simulation

4.3.1 Paramètres de simulation

Pour simuler notre solution, nous avons développé le modèle en Python. Les simulations ont été effectuées avec un ensemble de 513 données, comprenant 255 données normales et 258 données anormales. Le dataset a été divisé en deux parties : 80% pour l'entraînement du modèle et 20% pour le test.

4.3.2 Création du modèle de comportement

La création du modèle de comportement pour garantir la sécurité et la fiabilité des communications ferroviaires est un processus complexe, impliquant plusieurs étapes importantes. Tout d'abord, les données sont chargées depuis un fichier CSV, incluant des informations sur les horaires des trains, les arrêts, la vitesse maximale, les coordonnées

géographiques et les retards. Les heures de départ prévues et d'arrivée sont converties en minutes depuis minuit pour faciliter les calculs. Les coordonnées géographiques sont extraites des points géolocalisés, et les lignes contenant des valeurs manquantes sont supprimées pour garantir l'intégrité des données. Seules les colonnes pertinentes pour l'analyse sont conservées, et une nouvelle caractéristique, le temps de voyage, est calculée. Les données sont ensuite séparées en caractéristiques et cible, puis normalisées pour améliorer la performance des modèles de machine learning.

Un réseau neuronal récurrent (RNN) est utilisé pour modéliser les dépendances temporelles dans les données des trains. Les données normalisées sont restructurées pour être compatibles avec l'entrée d'un RNN. Un modèle RNN complexe est construit avec plusieurs couches LSTM, des couches de dropout pour éviter le surapprentissage et des couches de normalisation pour stabiliser les apprentissages. Le modèle RNN est ensuite entraîné sur les données de train et validé sur un ensemble de validation pour garantir une bonne généralisation. Les sorties du RNN sont utilisées comme entrée pour l'algorithme Isolation Forest, qui détecte les comportements anormaux. Une recherche randomisée d'hyperparamètres est effectuée pour optimiser les performances de l'Isolation Forest, qui identifie les anomalies dans les données de sortie du RNN.

La figure 4.7 représente l'entrainemnt du réseau neuronal récurrent.

```
# Entrainement du modèle RNN avec une architecture plus complexe
model_rnn = Sequential([
    LSTM(256, return_sequences=True, input_shape=(X_train.shape[1], 1)),
    Dropout(0.2),
    BatchNormalization(),
    LSTM(128, return_sequences=True),
    Dropout(0.2),
    BatchNormalization(),
    LSTM(64),
    Dropout(0.2),
    BatchNormalization(),
    Dense(64, activation='relu'),
    Dense(1)
])
optimizer = Adam(learning_rate=0.001)
model_rnn.compile(optimizer=optimizer, loss='mse')
model_rnn.fit(X_train_reshaped, y_train, epochs=100, batch_size=32, validation_split=0.2)
```

FIGURE 4.7 – Entrainement du RNN.

La figure 4.8 illustre la détection des comportements anormaux.

```
# Sauvegarder le modèle RNN
#model_rnn.save('model_rnn.h5')
model_rnn.save('model_rnn.keras')

# Utilisation de la sortie du modèle RNN pour Isolation Forest
output_rnn_train = model_rnn.predict(X_train_reshaped)
output_rnn_test = model_rnn.predict(X_test_reshaped)

# Optimisation des hyperparamètres de l'Isolation Forest avec RandomizedSearchCV
param_dist_iforest = {
    'n_estimators': [50, 100, 200],
    'max_samples': ['auto', 0.5, 0.75, 1.0],
    'contamination': [0.1, 0.2, 0.3],
    'random_state': [42]
}
random_search_iforest = RandomizedSearchCV(IsolationForest(), param_distributions=param_dist_iforest, sco
random_search_iforest.fit(output_rnn_train, y_train)
best_iforest = random_search_iforest.best_estimator_
```

FIGURE 4.8 – Entrainement de isolation forest.

Les anomalies détectées par l'Isolation Forest sont ensuite classifiées par un SVM. Les données identifiées comme normales par l'Isolation Forest sont utilisées pour entraîner

le SVM. Une recherche randomisée d'hyperparamètres est effectuée pour optimiser les performances du SVM. Le SVM est ensuite entraîné sur les données filtrées et évalué pour garantir une classification précise des comportements normaux et anormaux.

La figure 4.9 représente l'entrainement de SVM pour la prise de décision.

```
# Sauvegarder le modèle Isolation Forest
joblib.dump(best_iforest, 'best_iforest.pkl')

# Utilisation de la sortie de l'Isolation Forest pour SVM
anomalies_train = best_iforest.predict(output_rnn_train)
anomalies_test = best_iforest.predict(output_rnn_test)

# Filtrer les données normales pour l'entrainement du SVM
normal_train_indices = anomalies_train == 1
normal_test_indices = anomalies_test == 1

# Optimisation des hyperparamètres de SVM avec RandomizedSearchCV
param_dist_svm = {
    'C': [0.1, 1, 10, 100],
    'gamma': [1, 0.1, 0.01, 0.001],
    'kernel': ['rbf']
}
random_search_svm = RandomizedSearchCV(SVC(), param_distributions=param_dist_svm, refit=True, verbose=2,
random_search_svm.fit(output_rnn_train[normal_train_indices], y_train[normal_train_indices])
best_svm = random_search_svm.best_estimator_

# Sauvegarder le modèle SVM
joblib.dump(best_svm, 'best_svm.pkl')
```

FIGURE 4.9 – Entrainement de SVM.

4.3.3 Paramètre d'evaluation des performance

Dans cette partie, nous allons présenter les différentes métriques d'évaluation couramment utilisées en apprentissage automatique pour évaluer la performance de notre modèle de sécurité des communication dans le système TC-CBTC. Ces métrique nous permettent de mieux comprendre la précision et la capacité de notre modèle. Ces métriques incluent la précision, exactitude (Accuracy), le rappel (Recall) et le F1-score.

4.3.3.1 Matrice de confusion

La matrice de confusion est un outil essentiel pour évaluer les performances d'un modèle de classification. Elle permet de visualiser les performances du modèle en comparant les valeurs prédites aux valeurs réelles pour chaque classe. Cette matrice est indispensable pour définir les différentes métriques de classification telles que l'Accuracy, le F1-score,...

la figure 4.10 représente les éléments d'une matrice de confusion [26].

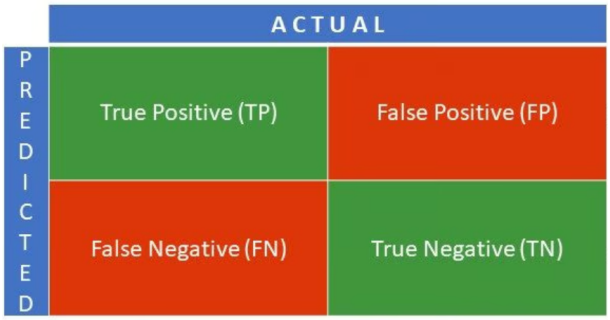

FIGURE 4.10 – Matrice de confusion.

Les éléments de la matrice sont disposés comme suit :
- Les lignes de la matrice représentent les classes réelles (ou vraies).
- Les colonnes de la matrice représentent les classes prédites par le modèle.

Les 4 cas possibles lors d'une prédiction binaire :
- Vrai négatif (True Negative) : est le nombre de cas où le modèle prédit correctement l'absence de la classe positive.
- Vrai positif (True Positive) : représente le nombre de cas où le modèle prédit correctement la présence de la classe positive.
- Faux positif (False positive) : est le nombre de cas où le modèle prédit incorrectement la présence de la classe positive.
- Faux négatif (False negative) : est le nombre de cas où le modèle prédit incorrectement l'absence de la classe positive.

4.3.3.2 Précision

La précision est une mesure de la performance d'un modèle de classification, spécifiquement dans le contexte des classes positives. Elle est définie comme la proportion de prédictions positives correctes parmi l'ensemble des prédictions positives effectuées par le modèle. En d'autres termes, la précision indique la fidélité du modèle lorsqu'il prédit une classe positive.

$$\text{Précision} = \frac{VP}{VP + FP} * 100 \tag{4.1}$$

4.3.3.3 Exactitude (Accuracy)

L'exactitude est définie comme la proportion des prédictions correctes (vrais positifs et vrais négatifs) parmi le nombre total de prédictions effectuées. Autrement dit, l'exactitude indique à quel point le modèle est bon pour identifier correctement les classes positives et

négatives. La formule pour calculer l'exactitude est la suivante :

$$\text{Exactitude} = \frac{VP + VN}{VP + FN + FP + VN} * 100 \tag{4.2}$$

4.3.3.4 Rappel (Recall) ou Sensibilité (Sensitivity)

Le taux de vrai positif (TVP), également appelé rappel ou sensibilité, est le pourcentage des exemples positifs qui sont correctement classés comme positifs. La formule pour calculer le Rappel est la suivante :

$$\text{Rappel} = \frac{\text{VP}}{\text{VP} + \text{FN}} * 100 \tag{4.3}$$

4.3.3.5 F1-score

Le F1-Score est une mesure de la performance d'un modèle de classification qui prend en compte à la fois la précision (precision) et le rappel (recall). Il est défini comme la moyenne harmonique de la précision et du rappel, ce qui en fait une métrique utile pour évaluer les modèles lorsqu'il existe un compromis entre ces deux mesures. La formule pour calculer le F1-score est la suivante :

$$F1 = \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \times 2 \tag{4.4}$$

4.4 Résultats obtenus

Cette section est dédiée à l'évaluation et à l'analyse des performances de notre solution proposée. Nous commençons par examiner la matrice de confusion. Ensuite, nous explorons les différentes métriques de performance precision, rappel, exactitude et F1-Score.

4.4.1 Matrice de confusion

La Matrice de confusion de notre modèle de classification est représentée sur la figure 4.11

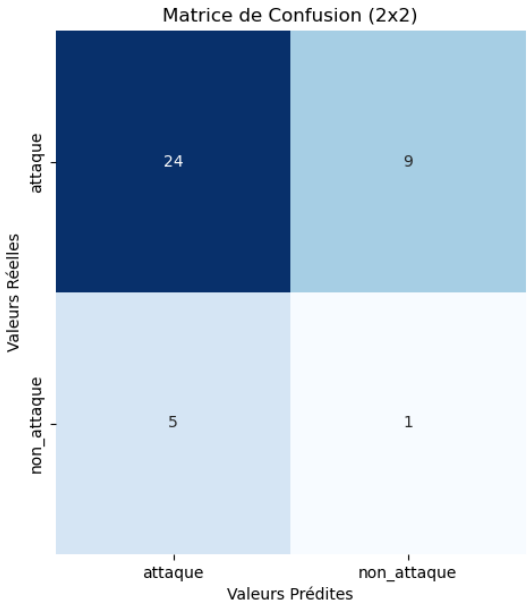

FIGURE 4.11 – La matrice de confusion de la solution proposée.

Avec 24 vrais positifs (VP), notre modèle a correctement identifié les cas positifs parmi ceux qui étaient réellement positifs. Cependant, les 9 faux positifs (FP) indiquent des erreurs où le modèle a prédit positif à tort. Ces prédictions erronées suggèrent une tendance à la sur-prédiction dans certaines conditions. Les 5 faux négatifs (FN) montrent les cas où le modèle a omis de prédire positif alors que la classe réelle était positive, bien que ce nombre soit relativement bas, ce qui indique une bonne sensibilité globale du modèle. Enfin, le seul vrai négatif (VN) souligne une prédiction correcte pour la classe négative parmi les cas réellement négatifs, bien que ce nombre soit limité.

Notre modèle présente une capacité raisonnable à détecter les occurrences positives réelles. Cependant, il nécessite une attention particulière pour réduire les faux positifs, ce qui pourrait améliorer encore sa précision et sa fiabilité dans des conditions plus diversifiées et complexes.

4.4.2 les performances obtenue

la figure 4.12 représente un histogramme qui determine les performances de notre modèle.
Nous avons :

— Vrais Positifs (VP) = 24
— Faux Positifs (FP) = 9
— Faux Négatifs (FN) = 5

— Vrais Négatifs (VN) = 1

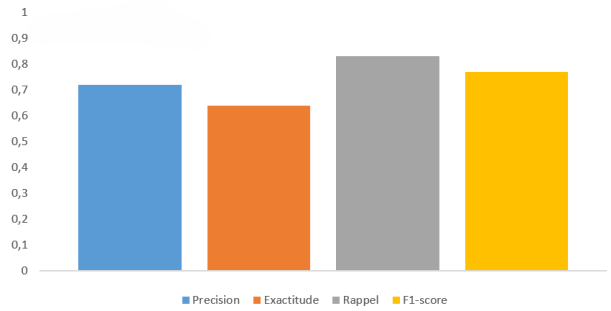

FIGURE 4.12 – Les performances de notre solution.

Dans cet histogramme, l'interprétation des différentes métriques est comme suit :

— **Precision :** 0.72, ce qui signifie que 72% des prédictions positives pour la classe attaque sont correctes. Cela indique que le modèle a une bonne capacité à identifier les attaques et que le modèle fait un nombre raisonnable de fausses alertes
— **Rappel :** 0.83, cela signifie que parmi toutes les instances réelles d'attaques, 83% ont été correctement identifiées comme des attaques par le modèle ; indique que le modèle est assez bon pour détecter la plupart des attaques.
— **Exactitude :** 0.64, ce qui indique que 64% des de toutes les prédictions (attaques et non-attaques) étaient correctes.
— **F1-score :** 0.77 ; un F1-Score de 77% montre que le modèle a un bon équilibre entre la précision et le rappel. Cela signifie qu'il est efficace pour détecter les attaques (haut rappel) tout en limitant le nombre de fausses alertes (bonne précision).

En conclusion, le modèle est performant dans un certain cadre, mais il nécessite des améliorations pour être considéré comme hautement performant et fiable.

4.5 Conclusion

Dans ce chapitre, nous avons évalué les performances de notre approche hybride pour les communications sécurisées dans le système TC-CBTC. Les résultats montrent que notre modèle est performant en termes de précision, rappel, exactitude et F1-Score. Cela démontre l'efficacité de notre méthode pour détecter et classifier les comportements des communications en temps réel. En conclusion, notre modèle est fiable et précis, ce qui le rend utile pour des applications de sécurité ferroviaire.

Conclusion Générale et Perspectives

Aujourd'hui très médiatisé, le système de transport ferroviaire intelligent est au cœur de toutes les attentions et représente un énorme défi pour les entreprises en termes d'innovation technologique. La sécurité ferroviaire est une préoccupation majeure dans notre société actuelle, et la reconnaissance des menaces potentielles ainsi que la sécurisation des communications peuvent jouer un rôle essentiel pour réduire les incidents et garantir le bon fonctionnement des infrastructures ferroviaires.

Dans ce contexte, nous avons proposé une approche hybride novatrice pour sécuriser les communications dans le système TC-CBTC (Train-Centric Communication-Based Train Control). Notre approche, intitulée HASC (Hybrid Approach for Secure Communications in TC-CBTC system), repose sur trois phases distinctes : l'authentification pour garantir l'identité des trains, la vérification de l'intégrité des données échangées, et la prise de décision pour déterminer les actions à entreprendre. Cette approche permet de prévenir les cyberattaques, les intrusions malveillantes, et les manipulations de données, assurant ainsi une communication sécurisée et fiable entre les trains.

Dans ce travail, nous avons donné une vue d'ensemble des systèmes de transport ferroviaire intelligents. Nous avons également inclus des définitions générales des méthodes sur lesquelles notre modèle repose. Nous avons ensuite approfondi notre compréhension de la sécurisation des communications ferroviaires en examinant les différentes méthodes proposées dans la littérature. Par la suite, nous avons détaillé notre approche hybride pour la sécurisation des communications ferroviaires. Enfin, en utilisant des datasets de données de communication ferroviaire très importants, nous avons évalué la performance de notre modèle sous le langage de programmation Python, en mesurant la précision, le rappel, l'exactitude et le F1-mesure, et les résultats sont très encourageants.

Notre modèle nécessite des données d'apprentissage de haute qualité et en grande quantité, ainsi que la prise en compte de facteurs contextuels pour une sécurisation plus précise des communications. Pour améliorer notre modèle, nous envisageons d'utiliser des techniques d'apprentissage plus avancées, d'augmenter la quantité et la qualité des données d'apprentissage et de prendre en compte davantage de facteurs contextuels.

Bibliographie

[1] Li, K., Wang, H., & Li, X. (2020). Intelligent Railway Systems : A Review. IEEE Transactions on Intelligent Transportation Systems, 21(5), 1806-1824. doi :10.1109/TITS.2019.2914446

[2] Corman, F., Meng, L. (2015). A Review of Online Dynamic Models and Algorithms for Railway Traffic Management. IEEE Transactions on Intelligent Transportation Systems, 16(3), 1276-1285. doi :10.1109/TITS.2015.2407996

[3] Tsanakas, P. (2017). Internet of Things in Railway Applications. Transportation Research Procedia, 25, 4485-4495. doi :10.1016/j.trpro.2017.05.345

[4] Hansen, I. A., Pachl, J. (2014). Railway Timetable & Traffic : Analysis, Modelling, Simulation. Eurailpress. ISBN 978-3-7771-0464-9

[5] Sussman, J. M. (2008). Perspectives on Intelligent Transportation Systems (ITS). Springer. doi :10.1007/978-0-387-74738-3

[6] Ma, J., Wang, H., Meng, L. (2017). Communication-Based Train Control (CBTC) Systems : A Comprehensive Review. IEEE Transactions on Intelligent Transportation Systems, 18(6), 1430-1449. doi :10.1109/TITS.2016.2615583

[7] Zhang, L., Lu, R., Wang, T., & Duan, Y. (2018). Design and implementation of CBTC system based on virtual coupling control strategy. In 2018 IEEE International Conference on Intelligent Transportation Systems (ITSC) (pp. 3801-3806). IEEE. doi :10.1109/ITSC.2018.8569801

[8] Xiaoxuan Wang, Lingjia Liu, Li Zhu, et Tao Tang. *Train-Centric CBTC Meets Age of Information in Train-to-Train Communications*. IEEE.

[9] International Electrotechnical Commission (IEC), IEC 62443-3-3 :2013 - Security for industrial automation and control systems - Part 3-3 : System security requirements and security levels, IEC, 2013.

[10] European Union Agency for Cybersecurity (ENISA), Railway Cybersecurity - Security measures in the railway domain, ENISA, 2020.

[11] Cortes, C., Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.

[12] OpenAI. *Datatron : A Platform for Model Management*. `https://www.openai.com/research/datatron`. Accessed : 2024-05-21.

[13] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.*

[14] Open Data Science Conference (ODSC). *ODSC - Open Data Science Conference*. `https://odsc.com/`. Accessed : 2024-05-21.

[15] Liu, F. T., Ting, K. M., Zhou, Z.-H. (2008). Isolation Forest. In 2008 Eighth IEEE International Conference on Data Mining (pp. 413-422). IEEE.

[16] MetalBlog. *La détection d'anomalies en machine learning non supervisé*. `https://metalblog.ctif.com/2022/10/03/la-detection-danomalies-en-machine-learning-non-supervise/`. Accessed : 2024-05-21.

[17] Lakshminarayana, S., Karachiwala, J. S., Chang, S.-Y., Revadigar, G., Kumar, S. L. S., Yau, D. K. Y., Hu, Y.-C. (2018). Signal jamming attacks against communication-based train control : Attack impact and countermeasure. In Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks (pp. 160-171).

[18] Gao, B., Bu, B. (2019). A novel intrusion detection method in train-ground communication system. IEEE Access, 7, 178726-178743.

[19] Song, Y., Bu, B., Zhu, L. (2020). A novel intrusion detection model using a fusion of network and device states for communication-based train control systems. Electronics, 9(1), 181.

[20] Falahati, A., Shafiee, E. (2022). Improve safety and security of intelligent railway transportation system based on balise using machine learning algorithm and fuzzy system. International Journal of Intelligent Transportation Systems Research, 20(1), 117-131.

[21] Lim, H. W., Temple, W. G., Tran, B. A. N., Chen, B., Kalbarczyk, Z., Zhou, J. (2019). Data integrity threats and countermeasures in railway spot transmission systems. ACM Transactions on Cyber-Physical Systems, 4(1), 1-26.

[22] Purwanto, Y., Ruriawan, M. F., Alamsyah, A., Wijaya, F. P., Husna, D. N., Kridanto, A., Nugroho, F., Fakhrudin, A., Itqon, M., Febrianta, M. Y., et al. (2023). Security architecture for secure train control and monitoring system. Sensors, 23(3), 1341.

[23] Wang, Y., Zhang, W., Wang, X., Guo, W., Khan, M. K., Fan, P. (2020). Improving the security of LTE-R for high-speed railway : from the access authentication view. IEEE Transactions on Intelligent Transportation Systems, 23(2), 1332-1346.

[24] Microsoft. *Microsoft Visual Studio Code*. Disponible sur : `https://code.visualstudio.com/`. 2024.

[25] Python Software Foundation. *Python*. Disponible à : `https://www.python.org/`.

[26] Geekflare, *Confusion Matrix in Machine Learning*, 2024. Disponible sur : `https://geekflare.com/fr/confusion-matrix-in-machine-learning/`. Accessed : 2024-06-02

Résumé

La sécurité des communications dans les systèmes ferroviaires, en particulier pour le TC-CBTC (Train-Centric Communication-based Train Control), est cruciale en raison de la complexité croissante et des menaces cybernétiques. Ce mémoire propose une approche hybride combinant apprentissage automatique et cryptographie pour garantir l'authentification et l'intégrité des données échangées. En utilisant des réseaux de neurones récurrents pour la détection proactive des anomalies, couplés à une méthode d'isolation forest et une machine à vecteurs de support, notre solution permet une classification précise des comportements anormaux. Les performances ont été évaluées à l'aide de diverses métriques, montrant l'efficacité et la robustesse de notre approche pour sécuriser les communications ferroviaires.

Mots clés : Systèmes de transport ferroviaire intelligents, Système TC-CBTC, Sécurité des communications, Apprentissage automatique, Cryptographie.

Abstract

The security of communications in railway systems, particularly for TC-CBTC (Train-Centric Communication-based Train Control), is crucial due to increasing complexity and cyber threats. This thesis proposes a hybrid approach combining machine learning and cryptography to ensure authentication and data integrity. By utilizing recurrent neural networks for proactive anomaly detection, coupled with an isolation forest method and a support vector machine, our solution enables precise classification of abnormal behaviors. The performance was evaluated using various metrics, demonstrating the effectiveness and robustness of our approach in securing railway communications.

Keywords : Intelligent Railway Transport Systems, TC-CBTC System, Communication Security, Machine Learning, Cryptography.