UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE
APPLIQUÉES

PAR
CHEIKH SOUHAIBOU AMAR

DÉVELOPPEMENT D'UN MODÈLE DE DÉTECTION D'ANOMALIE BASÉE SUR LES FORÊTS D'ISOLEMENT DANS UN ENVIRONNEMENT V2G

SEPTEMBRE 2022

Remerciements

Je remercie DIEU le tout puissant de m'avoir donné la santé et la volonté d'entamer et de terminer ce mémoire.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de mon directeur de recherche Mr Boucif Amar Bensaber, je le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant ma préparation de ce mémoire.

Ma gratitude va aussi aux professeurs Ismaïl Biskri et Mesfioui Mhamed qui, de par leurs minutieuses corrections, m'ont permis d'améliorer ce mémoire.

Je suis également très reconnaissant envers mes collègues du laboratoire « LAMIA » de l'université du Québec à Trois-Rivières qui ont collaboré à travers nos différents échanges à la réalisation de ce travail.

Enfin je remercie ma famille et tous ceux de près ou de loin m'ont soutenu et encouragé durant tout mon parcours scolaire.

Résumé

De par les réseaux intelligents, on assiste à une énorme exploitation efficace des sources d'énergie renouvelable grâce à un soutien technologique pour le transfert d'énergie entre les producteurs et les consommateurs d'énergie locaux.

Ce sont des systèmes cyber-physique complexes souvent avec des propriétés critiques pour la sécurité.

Avec l'augmentation du nombre de véhicules électriques (VE), la complexité des communications dans les réseaux véhiculaires électriques (V2G) a augmenté. Les Véhicules électriques participent non seulement au prélèvement d'énergie sur le réseau pour se charger, mais ils font également office de batteries qui peuvent se décharger sur le réseau pendant les périodes de forte demande d'énergie.
Le système Vehicle to Grid (V2G) introduit de nouveaux composants tels que les véhicules électriques, les agrégateurs et les équipements d'alimentation de véhicules électriques (EVSE) au réseau, dont chacun peut présenter des vulnérabilités supplémentaires ; ces composants introduisent également de nouvelles voies de communication qui soulèvent de nouveaux problèmes de coordination entre de multiples parties prenantes.
Assurer la sécurité des flux de communication entre ces entités est essentiel pour assurer une intégration sécurisée dans le réseau.
À cette fin, nous proposons dans le cadre de ce mémoire un moteur de détection d'anomalies utilisant la résolution de forêt d'isolement pour détecter les intrusions dans les réseaux V2G.
Notre modèle garantit une détection d'anomalie sur les cyber-messages gérant divers changements d'état et de contraintes de données et nécessitant une faible complexité temporelle linaire et une petite exigence de mémoire.

Pour former le modèle et évaluer les performances, nous avons utilisé une base de données V2G issue du simulateur MiniV2G comportant des scénarios d'attaques pour démontrer la faisabilité de la technique des forêts d'isolement.

Abstract

Through smart grids, there is a huge efficient exploitation of renewable energy sources through technological support for energy transfer between local energy producers and consumers.

They are complex cyber-physical systems often with security-critical properties.

With the increase in the number of electric vehicles (EVs), the complexity of communications in electric vehicular (V2G) networks has increased. EVs not only help draw energy from the grid to charge, they also act as batteries that can discharge to the grid during periods of high power demand.

The Vehicle to Grid (V2G) system introduces new components such as electric vehicles, aggregators, and electric vehicle power equipment (EVSE) to the grid, each of which may have additional vulnerabilities; these components also introduce new communication channels that raise new coordination issues among multiple stakeholders.

Ensuring the security of communication flows between these entities is essential to ensure secure integration into the network.

To this end, we propose in this thesis an anomaly detection engine using isolation forest resolution to detect intrusions in V2G networks.

Our model guarantees anomaly detection on cyber messages handling various state changes and data constraints and requiring low linear time complexity and small memory requirement.

To train the model and evaluate the performance, we used a V2G database from the MiniV2G simulator with attack scenarios to demonstrate the feasibility of the isolation forest technique.

Table des matières

Table des figures

Liste des tableaux

Nomenclature

V2G	Vehicle-To-Grid
V2I	Vehicle-To-Infrastructure
V2V	Vehicle-To-Vehicle
IDS	Intrusion detection System
EV	Electric Vehicle
EVCC	Electric Vehicle Communication Controller
EVSE	Electric Vehicle Supply Equipment
SECC	Supply Equipment Communication Controller
ISO	International Organization for Standardization
EXI	Effficient XML Interchange
XML	Extensible Markup Language
MITM	Man In The Middle
DOS	Denial of Service
CA	Certificate Authority
DPI	Deep Packet Inspection
SVM	Support Vector Machines
RNN	Recurrent Neural Network

Chapitre1 - Introduction général

La technologie V2G permet le transfert bidirectionnel d'énergie entre les véhicules électriques et le réseau. Cette technologie consiste à transformer chaque véhicule en un système de stockage d'énergie, à améliorer la fiabilité énergétique et la quantité d'énergie renouvelable disponible pour le réseau pendant les pics de consommation [1]. Il existe 50 projets V2G dans le monde entier, qui tentent de trouver un moyen de faire fonctionner cette technologie pour tous les acteurs : propriétaires de VE, services publics de VE, équipementiers automobiles et développeurs de systèmes de recharge de VE et les développeurs de systèmes de chargement de VE [2].

Les capacités de stockage des réseaux véhiculaires électriques V2G permettent aux véhicules électriques de stocker et de décharger l'électricité générée par des sources d'énergie renouvelables telles que l'énergie solaire et éolienne, avec une production fluctuant en fonction des conditions météorologiques et climatiques au moment de la journée. Le réseau véhiculaire électrique V2G est formé de véhicules pouvant être connectés à une prise électrique. Ceux-ci sont communément appelés véhicules électriques rechargeables (PEV) et comprennent les véhicules électriques à batterie (BEV) et les véhicules électriques hybrides rechargeables (PHEV). Étant donné que 95 % des voitures sont garées à tout moment, les batteries des véhicules électriques peuvent être utilisées pour fournir de l'électricité de la voiture au réseau et vice versa [2].

Le système V2G introduit de nouveaux composants tels que les véhicules électriques, les agrégateurs et les équipements électriques (EVSE), et donc de nouveaux canaux de communication qui posent de nouveaux défis pour la coordination entre les multiples parties prenantes. Par conséquent, ces systèmes peuvent présenter encore plus de vulnérabilités. Assurer la sécurité des flux de communication entre ces entités est essentielle pour garantir une intégration sécurisée dans le réseau.

Comme tout autre réseau, V2G est soumis aux risques de cybercriminalité, à cet égard plusieurs types d'attaques ont été exploitées dans le réseau V2G. Il s'agit notamment des attaques réseau et des attaques basées sur les composants.

Les attaques réseau incluent le déni de service (DOS) et les attaques de l'Homme du Milieu (MITM). Ces deux attaques peuvent être lancées lors de la connexion du véhicule au réseau ou lors de la mise à jour du firmware et cela conduit à affecter la stabilité, la sécurité et la fiabilité du système. Cela peut même affecter le fonctionnement des composants V2G. Ceux-ci incluent la violation d'authentification et/ou d'autorisation au niveau du système de distribution (composants du réseau électrique, y compris les agrégateurs), l'usurpation d'identité des composants du système de distribution, la compromission des appareils finaux (véhicules électriques). Ces menaces rendent le sous-système V2G très vulnérable aux attaques. Par conséquent, la sécurité du système V2G est d'une importance primordiale pour son fonctionnement.

Pour faire face à ces attaques ou en d'autres termes pour réduire l'impact de ces attaques plusieurs solutions ont été proposées et exploitées. Parmi ces solutions figurent les systèmes de détection d'intrusion (IDS), qui sont conçus pour surveiller les événements survenant dans un système d'information afin d'identifier les signes de problèmes de sécurité.

Différentes approches ont été adopté pour la détection des menaces. Ces approches peuvent être mises en œuvre à l'aide d'une variété d'algorithmes d'apprentissage automatique (ML) tels que la machine à vecteurs de support (SVM), l'analyse en composantes principales (PCA), les réseaux de neurones (NN), l'algorithme de sélection négative (NSA), etc.

Même si ces approches présentent des avantages tels que protéger les systèmes contre les menaces découvertes et favoriser le partage de règles dans la communauté et la capacité de détecter les attaques invisibles et nouvelles, elles présentent également des inconvénients, parce qu'elles ne permettent pas de protéger contre les attaques de type zero-day (attaques contre vulnérabilités qui n'ont pas encore été corrigées ou divulguées), et les signatures doivent être constamment mises à jour pour suivre les nouvelles menaces entrantes.

Pour certaines techniques de détection d'anomalies, si les instances normales dans les données de test n'ont pas suffisamment d'instances normales similaires dans les données d'apprentissage, le taux de faux positifs pour ces techniques est élevé. Aussi la performance de ces techniques dépend fortement du choix de la mesure théorique de l'information. De telles mesures ne peuvent souvent détecter des anomalies que lorsqu'il existe un nombre élevé d'anomalies présentes dans les données.

À cette fin, nous proposons dans ce mémoire les forêts d'isolement comme modèle de détection d'anomalies basé sur l'apprentissage non supervisé. Elles possèdent une faible complexité de calcul et une grande applicabilité aux données complexes et de grande dimension et peuvent être utilisés sur des ensembles de données mixtes contenant des variables continues et discrètes, ce qui facilite l'exploitation des données disponibles lors du développement du modèle. Elles peuvent être utilisées dans des programmes d'apprentissage semi-supervisés et non supervisés. Contrairement à la plupart des algorithmes de détection d'anomalies, nous nous sommes appuyés sur la mise en évidence des anomalies par le concept d'isolement, qui améliore la capacité à détecter les attaques. Une base de données du simulateur MiniV2G est utilisée pour la formation et l'évaluation de notre modèle.

Dans la suite de ce mémoire, le chapitre 2 présente le réseau V2G d’un point de vue global en spécifiant les différents acteurs ainsi que la norme ISO 15118. Le chapitre 3 permet d’introduire la notion de système de détection d’intrusion en détaillant ses différents composants. Le chapitre 4 consiste en une revue des travaux réalisés sur la sécurité du réseau véhiculaire électrique (V2G) et le réseau ad hoc véhiculaire (VANET). Dans le chapitre 5, nous démontrons le fonctionnement de la forêt d’isolement et nous fournissons la méthode pour construire les algorithmes des arbres et des forêts d’isolement. Dans la section 6, nous implémentons une session V2G grâce au simulateur MiniV2G et analysons l’efficacité de notre modèle. Enfin, nous terminerons par une conclusion générale.

Chapitre2 – Généralité sur le réseau V2G

2.1 Introduction

Dans ce chapitre, nous présentons les concepts généraux du réseau V2G, la norme ISO 15118 de même que l'ensemble des protocoles de communication spécifiés par cette norme.

2.2 Réseaux véhiculaires électriques (V2G)

Un certain nombre de facteurs poussent l'industrie automobile vers les véhicules électriques (VE), non seulement parce que les unités de contrôle électroniques (ECU) sont plus légères et capables de fournir plus de services, mais aussi parce que l'industrie a pris des mesures pour s'éloigner des combustibles fossiles des VE [3]. De plus, les progrès de la technologie des véhicules électriques ont changé les perspectives de l'industrie automobile. Les véhicules électriques sont considérés comme des candidats prometteurs pour remplacer les véhicules à carburant fossile. Non seulement ils ont le potentiel de conduire à un transport plus propre, mais ils peuvent également fournir des capacités de stockage d'énergie pour d'autres applications telles que V2G, Car-to-home, Car-to-load, V2I(Vehicle to Infrastructure) et V2V(Vehicle to Vehicle).

2.3 Norme ISO 15118

Au Québec, il existe 7000 bornes de recharge publiques, Bien que la plupart des recharges se fassent à peu près à la maison, l'afflux prévu de véhicules électriques pourrait ajouter du stress au réseau pendant les périodes de pointe entre 18 h et 21 h [2]. En 2010, la création de la norme ISO 15118 pour la communication V2G a été initiée. La norme internationale décrit les protocoles de communication numérique que les véhicules électriques et les bornes de recharge doivent utiliser pour charger les batteries des véhicules électriques. Cela inclut les applications de recharge filaires (CA et CC) et sans fil.

Les mécanismes de charge intelligents intégrés à la norme permettent au réseau d'adapter la capacité du réseau aux demandes énergétiques du nombre croissant de véhicules électriques connectés au réseau. L'ISO 15118 permet un transfert de puissance bidirectionnel pour les applications V2G en renvoyant la puissance du VE au réseau en cas de besoin. De plus, la

norme permet l'intégration des véhicules électriques dans le réseau intelligent. Un réseau intelligent est un réseau électrique qui repose sur les composants du réseau tels que les producteurs d'énergie, les consommateurs et les transformateurs par le biais des technologies de l'information et de la communication, comme le montre la figure 1 ci-dessous.

Figure 1: Vue d'ensemble simple de la communication entre les différents acteurs V2G [2]

De plus, la norme ISO 15118 permet aux véhicules électriques et aux bornes de recharge d'échanger des informations de manière dynamique et de négocier des horaires de recharge appropriés. Il est important que les véhicules électriques fonctionnent de manière compatible avec le réseau, ce qui signifie que le système de recharge gère la recharge simultanée de plusieurs véhicules tout en évitant la surcharge du réseau. Pour s'assurer que ces applications de recharge intelligente calculent un calendrier de recharge individuel pour chaque VE, elles utilisent les informations disponibles sur l'état du réseau, la demande d'énergie de chaque VE, ainsi que l'heure de départ et l'autonomie de chaque conducteur. Le système de charge est défini comme une communication entre l'équipement d'alimentation des véhicules électriques (EVSE) et le contrôleur de communication de l'équipement d'alimentation (SECC). Les communication V2G sont basées sur les définitions de la spécification ISO 15118, avec un total de huit spécifications de la norme. Les deux premières se concentrent sur différentes parties de la communication. Voir Figure 2 ci-dessous.

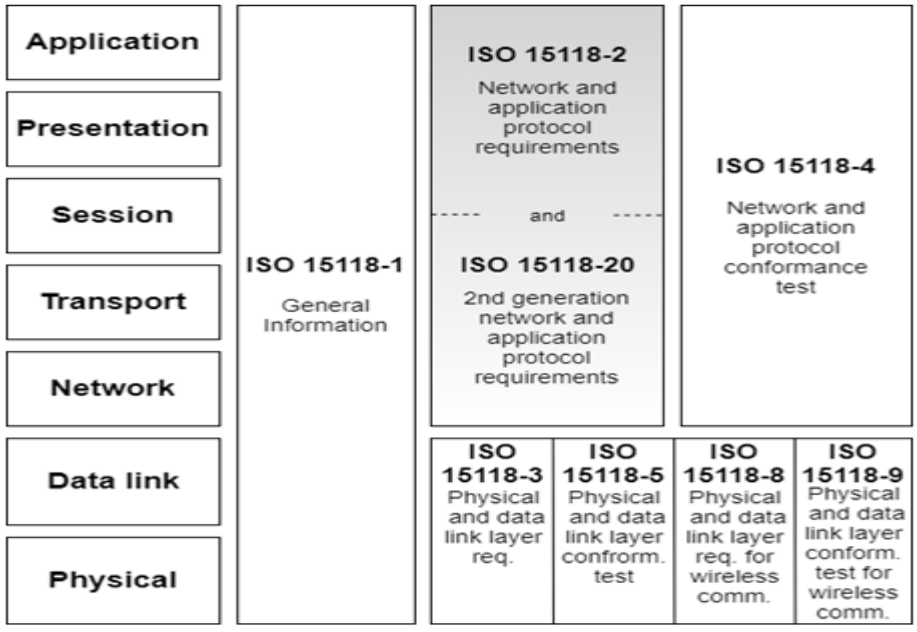

Figure 2: Les différentes spécifications de la norme ISO 15118. [4]

Dans ce mémoire, nous nous concentrons sur l'ISO 15118-2 : Exigences pour les protocoles de réseau et d'application. Cette spécification concerne la communication V2G. Elle comprend les différents types de messages et leur contenu respectif et les organigrammes qui ordonnent que les messages doivent parvenir au client ou au serveur.

2.4 Protocole de transport V2G (V2GTP)

Dans la spécification ISO 15118 [5], le protocole de transport V2G (V2GTP) est le protocole désigné pour le transfert de données V2G (pour la couche session) et peut être considéré comme HTTP pour la communication V2G. Un paquet V2GTP a un en-tête et un champ de charge. L'en-tête a une longueur de 8 octets et comporte quatre paramètres différents : la version du protocole, la version du protocole inverse, le type de message et la longueur de la charge utile, tandis que le champ de charge utile V2GTP renvoie les données réelles au transport. Quel que soit le type de message utilisé, un message V2G sera toujours encapsulé dans ce format. Une session démarre avec le client qui lance ses requêtes en multidiffusion, les requêtes dites SDP (SECC Discovery Protocol) sont envoyées via UDP. En raison de la paire requête-réponse SDP, un paramètre de sécurité est défini pour indiquer si la session

doit utiliser TCP ou TLS. Le serveur initie alors la configuration d'une session TCP ou TLS en négociant [5].

2.5 Représentation des données V2G

Les messages envoyés après une paire requête-réponse SDP utilise le format de représentation de données Efficient XML Interchange (EXI). XML/EXI est une représentation compacte du langage de balisage extensible (xml). En utilisant un algorithme relativement simple qui facilite des implémentations rapides et compactes, et un petit ensemble de représentations de types de données, EXI produit de manière fiable des encodages efficaces de flux d'événements XML [6]. Les communications V2G peuvent être spécifique au mode charge alternative (CA) ou au mode de charge continue (CC), et certains types de messages sont envoyés uniquement pour les communications CA ou CC correspondantes.

2.6 Modes d'identification

La spécification ISO 15118 a défini deux modes d'identification: le mode d'identification externe (EIM) et le mode d'identification Plug and Charge (PnC) [30]. Pour s'identifier et procéder la recharge dans le mode EIM, des équipements externes sont utilisés tel qu'une carte de paiement ou une carte RFID. Tandis que pour le mode PnC, les entités s'identifient d'une façon automatique à l'aide d'un contrat installé dans le véhicule. Pour assurer la sécurité des données échangées entre le véhicule et la borne dans ce dernier mode de charge, plusieurs mécanismes de cryptage sont déployés.

2.7 Sécurité dans la norme ISO 15118

Pour assurer la sécurité des communications au sein du réseau V2G, la norme ISO 15118 a proposé une infrastructure à clé publique (PKI). Aussi La spécification ISO 15118 convient à une approche hybride commune, utilisant un algorithme à clé asymétrique pour créer et vérifier des signatures numériques en plus de convenir d'une clé symétrique. Tous les messages peuvent ensuite être chiffrés et déchiffrés à l'aide de l'algorithme de clé symétrique pendant la session de chargement. Pour une session PnC réussie, le VE et le système de facturation doivent prendre en charge trois propriétés de sécurité principales : la

confidentialité, l'intégrité et l'authenticité. Tout d'abord, les messages doivent être cryptés pour garantir qu'aucun tiers ou acteur malveillant ne puisse espionner les communications. Deuxièmement, il est important que si le message ait été modifié, il soit détecté pour s'assurer de l'intégrité des données. Enfin, pour s'assurer que l'autre personne est bien celle qu'elle prétend être, les véhicules électriques et les systèmes de recharge doivent authentifier leurs homologues de communication. Pour s'assurer que ces trois fonctionnalités de sécurité soientt respectées, des sessions TLS sont utilisées pour sécuriser les communications V2G. Pour la sécurité de la couche de transport, les messages communiqués entre EVCC et SECC sont chiffrés à l'aide de la clé symétrique négociée lors de la phase de négociation de la clé TLS. Le protocole d'accord de clé pour accepter de partager une clé de session TLS symétrique est appelé Elliptic Curve Diffie-Hellman (ECDH). Ce dernier est un protocole d'accord de clé qui permet à deux parties, chacune avec une paire de clés privées/publiques (asymétriques), d'établir un secret partagé sur un canal non sécurisé. La clé symétrique est utilisée pour chiffrer et déchiffrer les messages à l'aide d'une suite de chiffrement par blocs nommés AES_128_CBC_SHA256. Pour déchiffrer le message de l'expéditeur, le destinataire a besoin d'une clé symétrique. Ainsi, sans accès à la clé, le message ne peut pas être déchiffré, ce qui garantit la confidentialité. Semblable aux suites de chiffrement TLS, la spécification ISO 15118 fournit un schéma d'accord de clé alternative pour établir des clés de chiffrement pour certains messages contenant des données sensibles. Cette clé de cryptage est utilisée pour assurer la sécurité des messages, principalement pour crypter les clés privées envoyées entre EVCC et un autre tiers. Dans ce schéma, le protocole ECDH statique éphémère est utilisé, ce qui permet la création d'une clé de session partagée à partir d'une clé publique statique qui existe déjà à l'intérieur du véhicule. Ceci est expliqué plus en détail dans l'annexe G de la spécification ISO 15118 [5]. La vérification de l'authenticité et de l'intégrité des données est une fonction effectuée par cryptographie asymétrique, à l'aide d'une paire de clés constituée d'une clé privée/publique. La clé privée doit être gardée secrète et utilisée uniquement par l'entité à laquelle elle appartient pour créer une signature numérique. Les clés publiques sont distribuées aux paires dans le même écosystème et sont utilisées pour vérifier les signatures créées avec la clé privée associée. La cryptographie à clé publique est utilisée pour créer et vérifier des signatures afin de vérifier l'authenticité de

l'expéditeur et l'intégrité du message reçu. L'algorithme de cryptage utilisé pour cela est l'algorithme de signature numérique à courbe elliptique (ECDSA). Outre les mesures de sécurité décrites dans la spécification ISO 15118, certaines méthodes de sécurité sont intégrées à TLS. Un exemple est l'algorithme de code de vérification de message, qui utilise des clés et des étiquettes symétriques pour vérifier efficacement l'authenticité d'un message.

La spécification ISO 15118 décrit un écosystème de certificats numériques que les PnC doivent posséder pour fonctionner correctement. C'est là qu'intervient l'infrastructure à clé publique (PKI). Une infrastructure à clé publique définit une infrastructure qui comprend différentes entités, politiques et dispositifs capables de gérer, de distribuer et de révoquer des certificats numériques basés sur un chiffrement asymétrique. Dans un système PKI, les clés publiques sont associées à des identités via le processus d'inscription et d'émission de certificats. La connexion est établie par une autorité de certification [CA]. Ces autorités de certification gèrent la création, le stockage, la distribution et la révocation des certificats numériques. Un certificat numérique est un document électronique utilisé pour vérifier qu'une clé publique appartient à une partie autorisée.

Par conséquent, il est également connu sous le nom de certificat de clé publique. Pour la spécification ISO 15118, il existe un certificat SECC signé par un opérateur de recharge certifié et déposé à la borne de recharge. L'EVCC les utilise pour authentifier SECC sur TLS. Il existe également des certificats contractuels dans l'EVCC pour l'authentification auprès du SECC et/ou des participants secondaires, qui sont utilisés pour signer la session V2G. Enfin, il existe des certificats racine V2G et des certificats de sous-CA potentiels pour attester des certificats SECC et des certificats contractuels associés. Tous ces éléments sont essentiels pour maintenir la confiance et la sécurité dans le processus PnC. L'une des plus grandes améliorations de la fonctionnalité PnC est que le conducteur n'a rien à faire pour authentifier le véhicule, il suffit de brancher le câble de charge dans le véhicule et la station de charge ; Pas besoin de saisir une carte de crédit, de scanner un code QR ou d'ouvrir une application. Cela rend la spécification ISO 15118 très conviviale.

2.8 Sécurité dans les réseaux véhiculaires électriques

La confidentialité garantie que les communications entre les parties autorisées sont confidentielles et qu'un accès non autorisé aux transmissions de données dans le véhicule ne soit pas possible. Il est également important de s'assurer que la modification non autorisée des données est impossible ou du moins détectable. La perspective d'une modification non autorisée des actifs du système ou des données réellement transmises, y compris l'écriture, la modification et la suppression de messages, est clairement mauvaise. Le but de l'authentification et de l'identification est d'établir et de s'assurer que la source d'un message est correctement identifiée. Dans les véhicules, cela est fortement recommandé car seuls les actifs authentifiés et identifiables peuvent communiquer avec un certain ECU. La sécurité des véhicules devient un problème sérieux dans l'industrie automobile, et des engagements ont également été pris d'un point de vue législatif (par exemple, le Spy Car Act américain). Alors que la conduite autonome et l'intégration des fonctions V2V et V2I gagnent du terrain, de nouvelles failles de sécurité sont inévitables. Cela oblige les constructeurs à utiliser des architectures de sécurité appropriées et éprouvées dans leurs véhicules. Lorsque l'on considère la sécurité des véhicules, l'un des aspects les plus importants est le souci de la sécurité des passagers.

Lorsque Miller et Valasek ont réussi à contrôler à distance un Jeep Cherokee 2014, ils ont montré qu'une panne du réseau interne pouvait avoir des conséquences fatales [7]. Fonctionnant à distance, ils envoient des messages au réseau de zone de contrôleur (CAN) pour activer les clignotants, verrouiller la voiture, contrôler la direction à distance et couper le moteur. Bien que la sécurité soit la priorité numéro un, les questions de confidentialité et d'intégrité doivent également être prises en compte, tout comme les entreprises modernes. À titre d'exemple pratique, personne ne devrait être en mesure de suivre un conducteur en accédant aux données GPS ou en écoutant un haut-parleur.

Le tableau 1 ci-dessous montre trois catégories possibles d'intrus de véhicules : les utilisateurs de première main, tels que les propriétaires de voitures ; les constructeurs automobiles, les mécaniciens automobiles et le personnel de garage ; et diverses parties non autorisées, telles que les agences.

Attaquants	Connaissance	Accès physique
Propriétaire	Varié (la plupart du temps faible)	Plein
OEM, mécaniciens, etc.	Haut	Plein
Tiers non autorisé	Varié (peut être élevé)	Limité ou nul

Tableau 1: Attaquants possibles ou intrus effectuant des modifications non- autorisé.

Les deux premiers groupes d'attaquants ont un accès physique complet à tous les supports de transmission et aux appareils correspondants concernés dans le réseau automobile, et le deuxième groupe représente le plus grand risque. Les personnes de ce groupe peuvent avoir des connaissances de base et des outils techniques suffisants pour mener à bien une attaque par intrusion, ce qui pourrait entraîner des dommages permanents au logiciel et au matériel du véhicule. Un motif possible d'attaques non autorisées par des tiers pourrait être d'obtenir des informations privées sur les passagers par le biais d'écoutes téléphoniques ou de vol de données.

2.9 Défis

La plupart des problèmes de sécurité prévisibles sont des sous-produits de l'interconnexion de divers systèmes de bus de véhicules. Le Réseau d'interconnexion local (LIN), Réseau de zone de contrôle (CAN) ou Transport de systèmes orientés médias (MOST) sont des systèmes de bus interconnectés qui peuvent accéder et envoyer des messages à n'importe quel autre calculateur. De plus, un seul système de bus compromis peut mettre en péril l'ensemble du réseau de communication embarqué. Combinées à l'intégration croissante de réseaux externes tels que V2V, V2G et V2I, les futures attaques contre les systèmes de communication automobile peuvent être réalisées sans aucun contact physique, simplement en marchant à côté d'une voiture ou avec un téléphone portable dans presque tous les endroits dans le monde [8].

2.9.1 Appareils légers

Les communications modernes sur Internet permettent des solutions de sécurité pour l'authentification, la vérification de l'intégrité et l'utilisation de protocoles et d'algorithmes basés sur le cryptage pour assurer la confidentialité. Dans de nombreux cas, ces propriétés de sécurité peuvent être fournies sans effort et elles introduisent une latence limitée avec la puissance de calcul moderne d'aujourd'hui. Cependant, considérons un réseau automobile moderne ; composé de plusieurs appareils plus petits (commutateurs, microcontrôleurs, ordinateurs, capteurs, etc.) qui fonctionnent à une puissance inférieure, avec des taux de puissance de traitement et de ressources mémoire disponibles inférieurs. Si des schémas de protocole cryptographique ou d'autres mesures de sécurité doivent être appliqués dans un environnement automobile, le choix des algorithmes utilisés doit être soigneusement étudié afin de ne pas affecter les performances de ces dispositifs légers.

2.9.2 Connectivité

On pourrait soutenir que la connectivité est une arme à double tranchant en matière de sécurité à bord du véhicule. Au fur et à mesure que les architectures d'un réseau automobile évoluent, de nouvelles capacités utiles seront disponibles avec l'introduction des concepts V2V, V2G et V2I. Par exemple, les véhicules pourront trouver des places de stationnement disponibles dans les parkings, partager des informations sur les dangers immédiats sur la route ou permettre aux systèmes de conduite autonome de fournir des informations sur le trafic. Avec tous ces avantages, il y aura un compromis entre la commodité et la surface d'attaque potentielle introduite pour l'exploitation ou l'invasion de la vie privée. Par conséquent, la sécurisation des composants et des sous-domaines individuels ainsi que de l'ensemble du réseau interne de la voiture est essentielle pour ce nouveau paradigme de connectivité.

2.10 Analyse de sécurité

2.10.1 Images de la menace

Le développement des infrastructures de recharge, des réseaux de recharge et la mise en place de nouveaux acteurs secondaires ouvrent la voie à de nouvelles surfaces d'attaque exploitées par les attaquants. Par conséquent, à mesure que l'adoption de la recharge V2G

augmente, la protection de cet écosystème en pleine croissance deviendra plus courante. Il existe plusieurs scénarios à prendre en compte lors de l'examen de la surface d'attaque d'une session V2G, nous en avons identifié quatre principaux sur lesquels il faut se concentrer du point de vue de la sécurité. Ces scénarios sont décrits dans le tableau 2.

Scénario	Description	Objectif d'attaque
EVCC malveillant	L'attaquant a le contrôle sur l'ECU EVCC qui exécute le logiciel client V2G et est capable de diriger une série d'attaques.	Serveur d'échappement, refuser le client de chargement, commencer les communications TCP ouvertes, DOS sur SECC.
SECC malveillant	L'attaquant a le contrôle sur le serveur SECC et est capable de diriger une série d'attaques.	Refuser le service EVCC, DOS de l'EVCC, affecte la qualité de l'énergie dans le système.
L'homme du milieu	L'attaquant a le contrôle sur un nœud intermédiaire et peut modifier, injecter, rejouer ou envoyer des messages à volonté.	DoS sur EVCC en modifiant les paquets, sniffer les informations, envoyer des informations de comptage incorrectes.
Attaquant distant	L'attaquant est sur le réseau de charge et peut renifler, rejouer ou envoyer des paquets.	DOS sur serveur ou client, renifler des informations, détourner une session.

Tableau 2: Scénarios d'attaques d'une communication V2G.

Lors d'une session V2G, le contrôleur de véhicule (EVCC) et le contrôleur de capacité de charge (SECC) échangent des informations via HomePlugGreenPhy (Physical Support for Smart Charging). Essentiellement, ils interagissent comme deux ordinateurs sur un réseau public, et des parties malveillantes peuvent exister sur ce réseau. Par conséquent, il est important que ces réseaux soient correctement sécurisés et maintiennent un haut niveau de sécurité.

2.10.2 Gestion des certificats

La première préoccupation majeure est le traitement des certificats. Si les certificats du système PKI sont divulgués ou si l'adversaire peut falsifier la signature, cela entraînera une

insécurité de l'ensemble de la communication V2G à plusieurs niveaux. Pour ne citer que deux exemples, un attaquant pourrait facturer au nom d'une autre personne (s'il possède un certificat de contrat), ou un attaquant pourrait mettre en place une borne de recharge malveillante (à l'aide d'un certificat SECC). La spécification tente implicitement d'atténuer ce problème en limitant la durée de vie des certificats de contrat (jusqu'à deux ans à compter de l'expiration), obligeant les clients à s'engager à envoyer des demandes Certificate-Update pour continuer à utiliser PnC.

En outre, les autorités de certification (AC) et les sous-AC doivent donc prendre les mesures appropriées pour contrôler la délivrance de nouveaux certificats et révoquer de manière appropriée tous les certificats compromis connus dans le système. En plus de cela, les équipementiers et autres parties intéressées doivent également protéger les clés privées stockées dans EVCC et SECC. Une option comprendrait idéalement l'utilisation d'un modèle de sécurité matérielle (HSM) pour un calcul cryptographique efficace et surtout la restriction de l'accès aux clés privées.

2.10.3 UDP en V2G

TLS est la base de toutes les propriétés de sécurité maintenues au niveau de la couche transport. Sans TLS, la communication serait ouverte et lisible, permettant l'exécution de nombreux vecteurs d'attaque. Par exemple, cela augmente la probabilité d'attaques de piratage TCP à distance via l'usurpation d'adresse IP, ce qui permet à un attaquant d'envoyer des messages qui semblent provenir d'un client EVCC. Par conséquent, les fonctionnalités de sécurité de TLS atténuent efficacement les inquiétudes concernant les attaques à distance et autres attaques similaires. Dans cette optique, la phase de configuration du protocole est intéressante car elle négocie les termes de TLS. Si vous modifiez les paramètres de sécurité de la demande (SDP), vous pouvez démarrer la communication sans TLS en rétrogradant le client pour utiliser une configuration non-TLS. Par conséquent, il est essentiel pour les fabricants d'équipements d'origine (OEM) d'exiger TLS lors du déploiement de versions personnalisées de la spécification ISO 15118. L’attaque de déni de service (DOS) peut être effectuée à l'aide de la fonction de multidiffusion UDP pendant la phase de démarrage du protocole. Cela peut être fait par un client EVCC malveillant ou par un attaquant distant qui

envoie une demande de portée SDP à SECC via UDP. En conséquence, un attaquant peut drainer la bande passante du SECC, empêchant la borne de recharge de continuer son service normalement. Pour éviter cela, le serveur SECC doit appliquer une protection de base aux demandes attendues de deux octets de long et ne pas répondre à plus de 50 SDP (comme défini dans le protocole) à partir d'une seule adresse IP.

2.10.4 TCP en V2G

Le protocole V2G définit dans la spécification ISO 15118 est un protocole déterministe dans lequel les messages ne sont envoyés qu'à certaines étapes du protocole. Du point de vue de la sécurité, ce comportement déterministe peut être considéré comme un avantage, puisque l'attaquant doit se conformer à la prochaine structure de message attendue si l'attaque n'est pas détectée. Un exemple pratique serait par exemple, qu'après avoir envoyé des requêtes SessionSetup, le serveur ne devrait plus répondre à ces requêtes. Cela réduit la surface d'attaque et empêche les comportements malveillants tels que le fuzzing, les faux messages et les attaques par inondation. Dans cet esprit, les OEM déployant ce protocole doivent appliquer strictement ces statuts et ces séquences de messages lorsqu'ils implémentent la norme ISO 15118.

Par conséquent, il est important que le nombre d'adresses associées et de ports ouverts soit limité et ne laisse aucune fuite d'information sur les analyses de port et n'accepte pas les tentatives de connexion aléatoires.

2.10.5 Sécurité XML et EXI

Le protocole V2G utilise des messages XML au format EXI compressés avec des attributs de sécurité XML pour la signature et le cryptage. Les signatures doivent être utilisées pour les opérations qui nécessitent des exigences d'authentification, d'autorisation et d'intégrité (c'est-à-dire, lors de l'autorisation de recharge d'un véhicule ou de la signature d'informations de comptage). D'autre part, le chiffrement n'est utilisé que pour masquer la clé privée transmise via les messages CertificateInstallation et CertificateUpdate. Notez que ce chiffrement se situe au niveau de la couche application et bénéficie toujours implicitement des propriétés cryptographiques de TLS. Le système hybride fonctionne ensemble pour

protéger les points clés d'échange d'informations et protéger les données sensibles envoyées au cours d'une session.

En ce qui concerne la manière dont la clé de cryptage est établie ; selon le certificat de contrat, la clé utilisée pour la signature existe déjà dans le véhicule. En utilisant la clé privée associée au certificat de contrat, la clé partagée est dérivée via ECDH statique éphémère pour le chiffrement. Si la clé privée est compromise par un adversaire, le secret partagé peut être recréé à partir de la clé privée EVCC. Cette clé partagée peut ensuite être utilisée pour afficher toutes les données de session passées et futures envoyées à l'aide de cette clé. De même, nous voyons un besoin supplémentaire de sécuriser les clés stockées dans les véhicules. Il est important de maintenir une entropie élevée dans le schéma lors de la signature de documents avec des certificats de contrat à l'aide de l'algorithme ECDSA. Les défis doivent tous être créés avec une source appropriée d'aléatoire. Ceci est mentionné plus en détail dans la spécification [5]. Le standard XML a ses avantages et est largement adopté aujourd'hui comme moyen de structuration, de décodage et d'encodage des données.

2.10.6 Gestion des erreurs

Un autre sujet à considérer est la façon dont les erreurs sont gérées au cours d'une session. Dans la spécification ISO 15118, les méthodes de gestion des erreurs au niveau de l'application sont clairement spécifiées, telles que la spécification du moment où un serveur ou un client doit envoyer certains codes d'erreur. Un aspect non couvert par la spécification est la façon dont EVCC et SECC doivent se comporter dans cette situation, qui dépend en grande partie de l'OEM et les fournisseurs se définissent eux-mêmes. Par exemple, un client peut être amené à quitter une session après un certain nombre d'erreurs, ce qui peut effectivement mettre fin à la session en provoquant une chaîne de telles erreurs. Un autre exemple pourrait être une configuration où l'envoi d'une erreur du côté serveur inciterait le client à envoyer une nouvelle demande.

Cela pourrait être exploité pour essayer de forcer le client dans une boucle infinie, incapable de quitter ou de continuer la session.

CONCLUSION

Dans cette section, nous avons introduit le fonctionnement du réseau V2G et de la norme ISO 15118 en spécifiant les différents acteurs qui interagissent au sein du réseau ainsi que leurs rôles. Nous avons aussi introduit l'enjeu sécuritaire en spécifiant les risques qui se révèlent être sensible sur le bon fonctionnement du réseau.

Dans le chapitre suivant, nous aborderons les systèmes de détection d'intrusion, en spécifiant les catégories de détection. Cette étape nous permet de comprendre leurs fonctionnements dans le cadre général.

Chapitre 3 - Généralité sur le système de détection d'intrusion : IDS

3.1 Introduction

La détection d'intrusion est une approche pratiquée dans tous les secteurs depuis les débuts de la cybersécurité et elle est maintenant appliquée aux communications dans les réseaux véhiculaires. Les systèmes de détection d'intrusion (IDS) sont souvent utilisés comme périphériques physiques dédiés sur le réseau. De plus, ils sont utilisés comme deuxième ligne de défense dans les architectures de sécurité réseau, souvent stratégiquement placés derrière des pares-feux pour détecter les attaques en cours. Un exemple de placement d'IDS dans une topologie de réseau est présenté à la Figure 3.

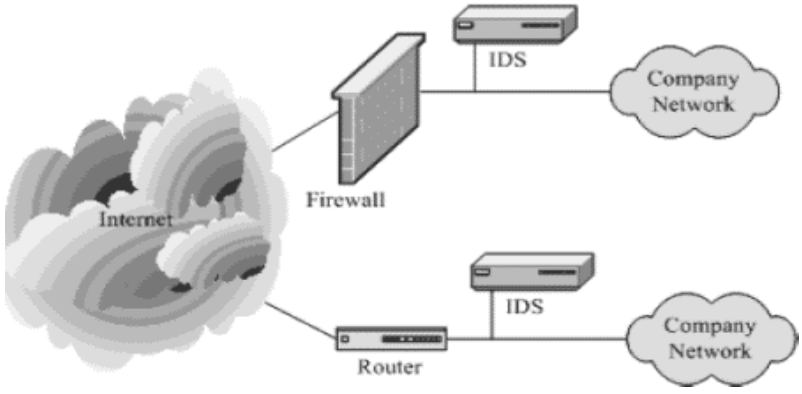

Figure 3: Placement d'un IDS à l'intérieur d'un réseau [9].

Contrairement aux pares-feux, dont le seul but est de filtrer le trafic entre les réseaux et d'empêcher les intrusions, les IDS surveillent et détectent les attaques ou les menaces qui existent au sein d'un réseau. L'objectif n'est pas d'empêcher les menaces malveillantes sur le réseau (à moins qu'il ne dispose de capacités de prévention), mais d'adopter une attitude plus proactive en alertant et en enregistrant les informations. L'IDS s'appuie généralement sur une inspection approfondie des paquets (DPI) étendue, qui peut filtrer les paquets au niveau de l'application. Les fonctions courantes d'un IDS incluent la mise à jour des systèmes pour les attaques futures, l'analyse des attaques potentielles sur le réseau, l'alerte d'autres mécanismes de sécurité existants ou la génération d'alertes pour les administrateurs réseau. Il

existe généralement deux méthodes d'IDS : le système de détection d'intrusion basé sur l'hôte (HIDS) et le système de détection d'intrusion basé sur le réseau (NIDS).

HIDS est un système IDS installé sur un hôte pour surveiller le comportement dynamique sur le système hôte. Il vérifie généralement les journaux d'audit du système et tente de détecter les activités malveillantes. Par exemple, HIDS peut vérifier l'intégrité de la vérification par rapport aux modèles de signature connus dans la base de données. NIDS, d'autre part, fait référence à un IDS qui réside sur un réseau. Il agit comme un renifleur en ligne, échantillonnant les paquets à mesure qu'ils arrivent sur le réseau. En règle générale, on voit des NIDS intégrés dans un commutateur réseau ou une passerelle centrale pour pouvoir gérer tout le trafic réseau. Par conséquent, les dispositifs NIDS doivent prendre en charge le traitement de paquets à grande vitesse pour éviter d'introduire des goulots d'étranglement dans le réseau.

Il existe différentes méthodes de détection pour détecter un comportement anormal dans l'IDS. À savoir les approches basées sur les signatures, les anomalies et les spécifications. Une façon consiste à utiliser l'une de ces fonctionnalités pour protéger le système, ou elles peuvent être utilisées uniformément comme IDS hybride.

3.2 Détection basée sur les signatures

Les moteurs de détection basés sur les signatures s'appuient sur la correspondance de modèles pour détecter les comportements malveillants connus dans des séquences d'octets ou des sous-ensembles d'instructions malveillantes. La mise en œuvre d'une approche basée sur les signatures nécessite de définir un ensemble unique de règles pour un réseau particulier. Par exemple, détecter et ne pas répondre aux requêtes ICMP ECHO permet à un hôte de ne pas être découvert par une analyse de scanneur de réseau (nmap). La commande «nmap» est utilisée pour découvrir des hôtes et des services sur un réseau informatique en envoyant des paquets et en analysant les réponses. Elle peut être utilisé pour effectuer des analyses de port ou des attaques par analyse de port.

Dans l'approche basée sur les signatures, l'algorithme de correspondance de modèles est le composant central qui détermine les performances. Étant donné que le logiciel IDS

fonctionne dans les limites d'une application typique, il existe des problèmes d'utilisation du processeur, de mémoire et de consommation d'énergie. Il y a eu quelques études sur des algorithmes améliorés par exemple l'algorithme de Myers [10] et Wu-Manber [11]. Cependant, l'algorithme à appliquer dépend des exigences du réseau et du matériel accessible sur un système donné. Un avantage distinct d'une implémentation basée sur les signatures est la possibilité d'utiliser plusieurs modèles de signature pour empêcher les modèles d'attaque connus, protéger le système contre les menaces découvertes et encourager les règles partagées dans la communauté. Bien que l'approche basée sur les signatures ait également ses inconvénients, principalement parce qu'elle ne peut pas protéger contre les attaques zero-day (attaques contre des vulnérabilités qui n'ont pas encore été corrigées ou divulguées), et les signatures doivent être constamment mises à jour pour suivre les nouvelles menaces entrantes.

3.3 Détection basée sur les spécifications

Une autre méthode de détection dans IDS est une approche basée sur les spécifications, où les écarts par rapport aux exigences de spécification définies sont utilisés pour détecter les paquets malveillants. Généralement, cela est réalisé en configurant des règles explicites pour chaque spécification de protocole afin de repérer les écarts par rapport au comportement normal. Des propriétés telles que la correspondance source et destination, le volume de données, l'heure du message ou les valeurs d'en-tête sont des exemples de métriques qui peuvent être prises en compte.

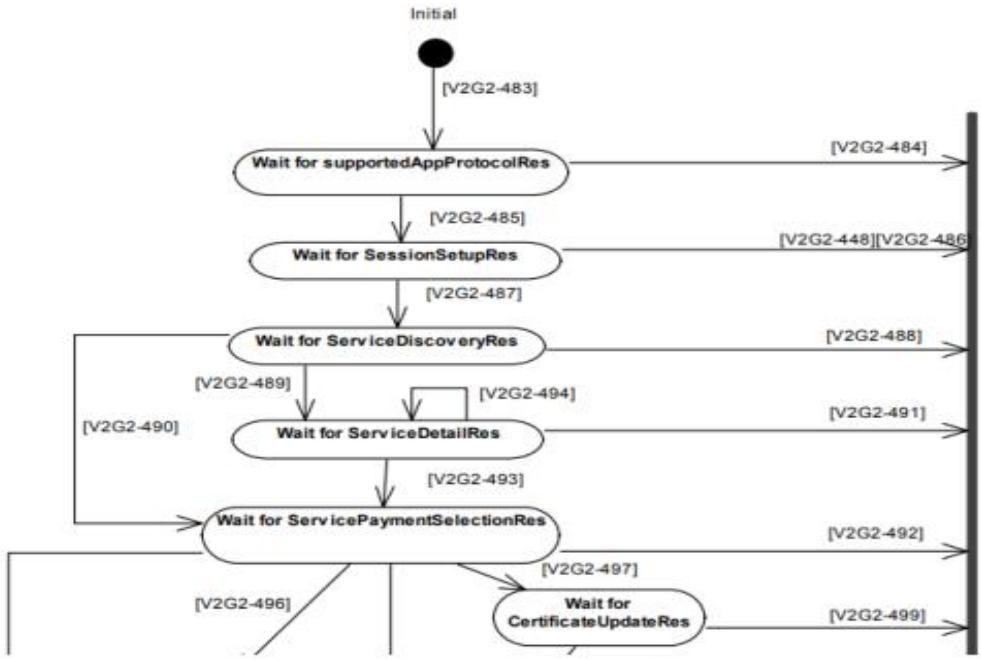

Figure 4: Extrait de communication V2G de la spécification ISO 15118

Contrairement aux autres méthodes, elle ne souffre pas de taux élevés de faux positifs. Cela s'explique en partie par le fait que ses principes basés sur des règles produisent des performances déterministes similaires aux méthodes basées sur les signatures. Par exemple, la norme ISO 15118 spécifie un organigramme dans lequel les utilisateurs peuvent suivre les types exacts de messages qui doivent être suivis les uns des autres. Par conséquent, un IDS basé sur les spécifications peut contenir une sorte d'arbre de décision pour garantir que la communication V2G respecte ses spécifications.

3.4 Détection basée sur les anomalies

Les approches basées sur les anomalies adoptent une approche différente pour détecter les menaces, mais détectent plutôt le trafic malveillant en s'écartant du comportement normal. En revanche, à partir des approches basées sur les signatures, l'IDS basé sur les anomalies tente de détecter des attaques jusque-là inconnues et de fournir une protection contre de futurs vecteurs d'attaque. Pour les méthodes statistiques, l'objectif est de trouver une base de caractéristiques de communication sur le réseau, telles que les types de paquets envoyés, la fréquence des paquets, la taille de la charge utile ou les modifications du contenu des

données et de la séquence des paquets. Une fois la ligne de base atteinte, l'IDS classe le trafic comme normal ou anormal et envoie une alerte si le système reçoit des paquets suspects. Par exemple, un système basé sur les anomalies peut détecter si une attaque par inondation ou par relecture est en cours s'il suit la fréquence à laquelle un type de message est généralement envoyé sur le réseau. Il peut également vérifier si des paquets sont supprimés en fonction des mêmes contrôles de fréquence. Comme mentionné précédemment, les méthodes basées sur les anomalies tentent de classer les lignes de base et de trouver des valeurs aberrantes pour les nouvelles menaces inconnues, des défis qui peuvent également être résolus en utilisant l'apprentissage automatique.

3.4.1 Approche basée sur la densité et la distance

La plupart des méthodes existantes de détection d'anomalies utilisent des méthodes de base de densité et de distance qui sont les suivantes :

- Réseau de neurones réplicateurs (RNN) : Il s'agit d'un réseau de neurones avec un perceptron multicouche à anticipation qui entraîne le réseau à répliquer relativement bien à des instances normales invisibles. Dans les RNN, ceux qui sont mal reconstruits sont considérés comme des anomalies.
- Machine à Vecteur de Support (SVM) à classe unique : La méthode trouve la plus petite région contenant la plupart des points de données normaux, les points en dehors de cette région sont considérés comme des anomalies.
- Une approche basée sur le regroupement construit des profils d'instances normaux, puis identifie les anomalies comme celles qui ne sont pas conformes aux profils normaux.

Pour les mesures de densité, "les points normaux apparaissent dans les zones denses, et les anomalies apparaissent dans les zones clairsemées."

Exemple : Les méthodes RNN et SVM regroupés pour les mesures de distance, "un point normal est proche de ses voisins, tandis qu'un point aberrant est éloigné de ses voisins".

Cependant, leurs capacités de détection d'anomalies ont souvent un « effet secondaire » ou un sous-produit d'algorithmes conçus à l'origine à des fins autres que la détection d'anomalies, ce qui entraîne les inconvénients suivants :

a) Trop de faux positifs c'est à dire identifier les instances normales comme des anomalies ou détecter trop peu d'anomalies.

b) De nombreuses méthodes existantes sont limitées à des données de faible dimension et de petite taille en raison des problèmes hérités de leur algorithme d'origine.

De plus, la détection d'anomalies de mesure de densité et de distance suppose qu'il y a des violations, par exemple une densité élevée et une courte distance n'impliquent pas toujours des cas normaux ; de même une faible densité et une longue distance n'impliquent pas toujours des anomalies. Pour surmonter l'inconvénient de détecter des anomalies dans les mesures de densité et de distance, nous allons utiliser dans le cadre de notre recherche la méthode de détection d'anomalie basée sur la forêt d'isolement.

CONCLUSION

Dans ce chapitre, nous avons présenté de façon détaillée les catégories des systèmes de détection d'intrusion ainsi que leurs fonctionnements. Chacune de ces méthodes possèdent des avantages ainsi que des limites. Cette étape nous permet de mieux comprendre le choix de notre modèle pour sécuriser le réseau V2G. Dans le chapitre suivant, nous présenterons l'état de l'art en spécifiant différents travaux effectués par des chercheurs sur la sécurité du réseau V2G ainsi que du réseau VANET.

Chapitre 4 – Revue de littérature

4.1 Introduction

En raison de l'importance des informations échangées au sein du réseau du véhicule au réseau (V2G), telles que les informations sur l'utilisateur, le contrat et les informations sur le véhicule. Ce réseau devient une cible importante pour les attaquants, ce qui le rend plus vulnérable.

Pour prévenir ces attaques et réduire leur impact, plusieurs travaux ont été menés, notamment des systèmes de détection d'intrusion (IDS) et des systèmes cryptographiques. Étant donné que le réseau V2G est une extension du réseau ad hoc véhiculaire (VANET) et que la littérature V2G n'a pas été profondément impliquée dans l'IDS, nous proposerons dans cette section quelques travaux proposant la sécurité des communications dans le V2G et le réseau VANET.

4.2 Sécurité au sein du réseau V2G

La plupart des travaux antérieurs sur la détection d'attaques ont utilisé des données de processus normales à l'aide de plusieurs algorithmes Machine learning (ML), tels que l'algorithme de sélection négative (NSA) [12], décomposition des valeurs singulières (SVD) [13], le réseau de neurones standard (NN) [14], Convolutional Neural Networks (CNN) [15], Recurrent Neural Networks (RNN) [16] et générative Adversarial Networks (GANs) mis en œuvre à l'aide de la méthode de mémoire à long court terme (LSTM) [17]. Ils reposent sur la construction de modèles capables d'analyser le comportement normal d'un système, puis d'identifier les observations non conformes comme des anomalies.

De plus, à mesure que la complexité, l'échelle et la non-linéarité du système augmentent, le développement de modèles de système haute-fidélité devient plus difficile. Les méthodes proposées dans [12][13][14], [16] peuvent souffrir d'un taux élevé de fausses alarmes et de mauvaises performances sur des données de grande dimension. De plus, certaines méthodes ont un coût de calcul élevé, telles que [15][16][17], tandis que d'autres, telles que [12], [13]

n'utilisent pas le processus en ignorant les informations sur les signaux de l'actionneur, ce qui peut contenir des informations précieuses sur l'état du processus.

En effet, la complexité de calcul des différents algorithmes d'apprentissage automatique varie, car les CNN et les RNN sont connus pour impliquer une grande quantité de calculs pendant les phases de formation et d'évaluation, tandis que les NN sont moins exigeants et vont d'un calcul moyen à élever [18]. En revanche, les algorithmes ML standard tels que Singular Value Decomposition (SVD), Support Vector Machine (SVM) se caractérisent par une complexité de calcul faible à moyenne, en fonction de la taille du problème.

En outre Z. Yang et al. [19] ont identifié les problèmes de confidentialité associés aux interactions V2G et ont proposé un mécanisme d'incitation basé sur la récompense pour les VE travaillant en tant que fournisseur de services. Dans cette méthode, les véhicules électriques utilisent des permis générés avec une méthode de signature numérique partiellement aveugle basée sur l'ID. Le permis est utilisé pour les transactions anonymes avec un agrégateur local. Il est suggéré que les véhicules électriques reçoivent les paiements en espèces électroniques qui devraient être utilisés en échange des services d'entretien du véhicule. Cette méthode vise à assurer la confidentialité d'un VE en tant que fournisseur de services lorsqu'il fournit des services auxiliaires au réseau. Cependant, cette méthode n'aborde pas la confidentialité d'un VE en tant que consommateur par rapport au service public d'énergie.

H. Li et al. [20] proposent un protocole d'authentification anonyme pour assurer la confidentialité des véhicules électriques et génère des reçus anonymes pour les transactions énergétiques. Un véhicule électrique collecte une signature aveugle sur les jetons du service public d'électricité et génère des pseudonymes pour chaque jeton. Pour établir des clés pour les pseudonymes, cette méthode nécessite que les EV utilisent un anonymiseur à latence élevée tel que The Onion Router (TOR). De plus, pour générer des reçus anonymes, les véhicules électriques doivent utiliser des signatures de groupe ou un réseau anonyme (par exemple, TOR). Cette méthode minimise la surcharge de calcul des véhicules électriques et permet à ces derniers d'obtenir des reçus anonymes pour les rapports qu'ils soumettent Cependant, ce mécanisme exige des changements importants dans l'infrastructure de

communication, ou des opérations gourmandes en temps de calcul par les véhicules électriques.

En outre, une nouvelle architecture de sécurité complète pour l'infrastructure V2G a été proposée par Vaidya B et al. [21]. Basée sur un modèle hybride d'infrastructure à clé publique (PKI), cette architecture intègre une certification croisée hiérarchique et des certifications croisées peer-to-peer. Cependant, ces études de recherche supposent que les propriétaires de véhicules font confiance aux agrégateurs pour qu'ils n'abusent pas des informations collectées ou ne les divulguent pas à des tiers. De plus, dans leur protocole, chaque EV est équipé d'une paire de clés publique/privée et du certificat de clé publique correspondant. En effet, le principal avantage de cette méthode est l'utilisation de la méthode de certification croisée peer-to-peer, dans lequel chaque CA de domaine est autonome dans la génération et la révocation des certifications croisées. Mais, ils ne sont pas très efficaces en termes de communication, car le certificat doit être transmis avec le message d'accès de chaque EV.

Liu et al. [22] ont conçu un système de préservation de la vie privée dépendant du rôle appelé ROPS. Il facilite les interactions sécurisées entre un VE et le réseau électrique et prend en charge les opérations de décharge centralisées et distribuées pour transférer l'énergie des VE vers le réseau. Cependant, comme leur méthode utilise la technique de signature en anneau, la charge de calcul sur un EV et la force d'anonymat dépendent de la taille du groupe d'agrégateurs. Ainsi, il existe un compromis entre les frais généraux et la force de l'anonymat. Plus important encore, en raison de l'utilisation de signatures aveugles, leur méthode ne parvient pas à responsabiliser les utilisateurs pour les comportements malveillants.

Un mécanisme de préservation de la vie privée pour les communications V2G a été conçu par Wan Z et al. [23], ce mécanisme peut assurer l'anonymat et l'intraçabilité des véhicules électriques dans l'authentification ainsi que la récompense de service. Cependant, cette méthode ne tient pas compte de la responsabilité et repose sur le tiers de confiance.

B. Vaidya et al. [24] proposent un mécanisme d'authentification décentralisé pour les véhicules électriques rechargeables (PEV) afin de résoudre les problèmes associés aux systèmes d'authentification centralisés. Cependant, le modèle contradictoire considéré dans cet article inclut des attaquants externes. La méthode proposée ne traite pas de la confidentialité vis-à-vis des entités internes ou des fournisseurs de services.

4.3 Sécurité au sein du réseau VANET

Les réseaux VANET ont beaucoup de choses communs sur la sécurité avec les réseaux V2G.

R. Kolandaisamy et al. [25] ont proposé la méthode d'analyse de flux multivariant (MVSA) pour détecter et atténuer les attaques DDoS sur VANET à l'aide de la simulation NS2. La méthode MVSA fournit une communication V2V via Restricted Stock Units (RSU), en déterminant un taux de charge utile moyen, une fréquence à différents moments et la durée de vie par véhicule pour chaque classe de grève. La méthode MVSA inspecte les fichiers de trace pour identifier le DDoS. Ensuite, MVSA décide du poids du flux, qui est suivi de la classification des paquets de flux comme légitimes ou malveillants. Cependant, alors que la méthode MVSA a démontré une stabilité et de bonnes performances, son inconvénient est que la réduction du délai des paquets n'est pas assurée pour détecter le nœud malveillant.

M. Raya et al. [26] proposent une technique qui consiste à réaliser un schéma de données malveillantes (MDS), un schéma de révocation de certificat en deux parties (RTC et RC2RL) et un schéma de vote à la majorité (LEAVE). Le MDS vise à détecter les fausses données dans les VANET et fonctionne en évaluant les données sensorielles, les messages reçus de ses voisins et un ensemble de règles d'évaluation. Cependant aucune règle d'évaluation spécifique n'est fournie.

Hasrouny et al. [27] ont démontré une technique de prédiction d'attaque améliorée. Cette technique peut prédire plusieurs types d'attaques VANET. En raison de sa méthodologie complexe, cette approche entraîne des frais généraux potentiels. Ainsi, elle n'est pas aussi efficace pour les applications en temps réel.

Al-Terri et al. [28] ont conçu deux approches collaboratives, à savoir la réputation de groupe (GR) et la détection coopérative (CD). Les deux techniques ont la capacité de détecter les nœuds malveillants au niveau de la couche MAC dans les VANET. Les deux approches surpassent les techniques disponibles pour détecter uniquement les attaques par déni de service distribué (DDOS). Cependant, les performances sont médiocres, en particulier en cas de détection d'attaque de trou de ver et de trou gris.

Yi Zeng et al. [29] ont proposé un modèle basé sur Deep learning in Vehicular Communication Module (DeepVCM): qui utilise une méthode de détection d'intrusion basée sur l'apprentissage profond dans VANET. Ils implémentent un système de détection d'intrusion pour le modèle de communication véhiculaire à l'aide d'algorithmes d'apprentissage en profondeur. Pour l'extraction de caractéristiques, l'algorithme CNN (Convolutional Neural Network) est utilisé et pour la classification, l'algorithme LSTM (Long Short-Term Memory) est utilisé. En effet cette technique possède un coup de calcul très élevé.

Safi et al. [30] ont conçu un protocole de communication véhiculaire sécurisé de bout en bout qui permet uniquement aux véhicules authentiques de transmettre les données entre les véhicules. Ainsi, il empêche les véhicules non autorisés de communiquer avec des appareils et des véhicules authentifiés. Cependant, cette technique échoue chaque fois qu'un type d'attaque se produit dans les VANET.

Zaid et al. [31] ont mis en place un système de détection d'intrusion (IDS) pour les VANET. L'IDS peut être déterminé en utilisant l'existence de nœuds indésirables (RN) qui peuvent lancer plusieurs attaques VANET. L'approche conçue a la capacité de surveiller une attaque de fausses données en considérant efficacement les approches statistiques, par contre cette technique n’arrive pas à détecter les attaques par dénis de services.

Les méthodes ci-dessus nécessitent d'apprendre un seuil commun et fiable pour déclencher des alertes d'anomalie, ce qui peut être difficile pour les réseaux d'information complexes. Pour remédier aux limites des travaux antérieurs, nous appliquons des forêts d'isolement avec une faible complexité de calcul et une grande applicabilité aux données complexes de

grande dimension. Ils peuvent être utilisés avec des ensembles de données mixtes contenant des variables continues et discrètes, ce qui facilite l'exploitation des données disponibles lors du développement de modèles. Ils peuvent être utilisés dans des programmes d'apprentissage semi-supervisés et non-supervisés. Contrairement à la plupart des travaux antérieurs, nous mettrons en évidence des anomalies basées sur le concept d'isolement, améliorant ainsi la capacité de détection des attaques.

Conclusion

Dans ce chapitre nous avons présenté quelques travaux portant sur la sécurité dans les réseaux V2G et VANET. L'étude de ces travaux montre qu'il existe différentes méthodes pour sécuriser les communications au sein des réseaux véhiculaire. Même si ces approches présentent des avantages tels que protéger les systèmes contre les menaces découvertes et favoriser le partage de règles et la capacité de détecter les attaques invisibles et nouvelles, elles présentent également des inconvénients. Pour certaines techniques si les instances normales dans les données de test n'ont pas suffisamment d'instances normales similaires dans les données d'apprentissage, le taux de faux positifs pour ces techniques est élevé. Aussi la performance de ces techniques dépend fortement du choix de la mesure théorique de l'information.

Dans le chapitre suivant, nous présenterons notre modèle de sécurité afin de combler les limites des travaux antérieurs sur la détection d'intrusion dans le réseau V2G. Nous allons introduire la notion de forêt d'isolement pour ensuite proposer des algorithmes pour la construction de ces dernières.

Chapitre 5 – Méthode de détection d'anomalie basée sur les Forêts d'isolements

5.1 Introduction

Dans ce chapitre, nous allons définir la notion d'isolement, pour ensuite spécifier les points sur lesquels est basé notre modèle, en spécifiant les différents algorithmes proposés ainsi que la méthode de prédiction du score.

5.1.1 Principe d'isolement

L'approche basée sur la Forêt d'isolement (iForest) est une méthode non Supervisé établi par Liu et al. [32]. Ces auteurs partent du principe d'Isolement sans utiliser aucune mesure de similarité ou de distance entre les instances. Pour réaliser cet isolement, l'auteur utilise un ensemble d'arbres d'isolement (iTree) effectué collectivement pour séparer les instances sur des sous-échantillons aléatoires des attributs et des instances. Dans le vocabulaire utilisé sur les arbres isolants, on appelle le chemin parcouru par l'instance x la séquence de nœuds dans l'arbre à traverser de la racine du point vers la terminaison du nœud contenant cette instance. Dans le cas d'une forêt d'isolement, on parle de la longueur moyenne du chemin à partir d'un point x, qui est simplement la moyenne des longueurs de chemin associées à ce point dans chaque chemin d'arbre de la forêt d'isolement.

En effet, la stratégie utilisée par iForest repose sur la supposition que dans une instance anormale, il y a quelques attributs dont la valeur est très différente des instances normales : un split choisi au hasard sur un tel attribut a donc beaucoup plus de chances de séparer une instance anormale des instances normales.

5.2 Notation

Supposons que nous avons à notre disposition un jeu de données X de n observations (instances) qui sont indépendantes, x_i , i = 1,...,n .

Chaque observation représente en effet un vecteur de dimension p possédant les valeurs de p variables : $x^{\mathcal{T}}_i = (x_{i1}, \ldots, x_{ip}),\ i = 1, \ldots, n$.

5.3 Notion d'Arbre d'isolement et Algorithme Proposé

L'idée d'isolation est réalisée par un partitionnement d'arbres binaires. Afin de construire un arbre, l'algorithme va choisir au hasard un échantillon $\mathcal{E}_u$ suivant les u observations de la base de données. L'arbre binaire de partitionnement est obtenu par une division de ce nouvel échantillon $\mathcal{E}_u$. Au début les instances de $\mathcal{E}$ sont dans la racine de l'arbre. Chaque séparation contient exactement deux nœuds qui sont représentées par un nouvel étage plus bas sur l'arbre. L'échantillon va être divisé en deux sous-groupes en choisissant de façon aléatoire une variable q de l'ensemble des p variables et un seuil de "split" s_q au hasard entre les valeurs comprises entre le maximum et minimum de cet attribut. Les instances ayant la valeur d'attribut inférieure à la valeur de séparation s_q partent à gauche et les autres à droite.

Soient : $\{x_i, i \in \mathcal{E}_u \mid x_q < s_q\}$ et $\{x_i, i \in \mathcal{E}_u \mid x_q \geq s_q\}$

Les deux sous-groupes obtenus, le processus est répété jusqu'à ce que toutes les instances soient isolées dans une feuille de l'arbre.

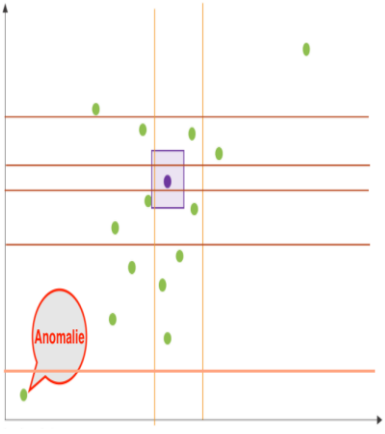

Figure 5: Représentation en deux dimensions d'un jeu de donnée contenant une anomalie

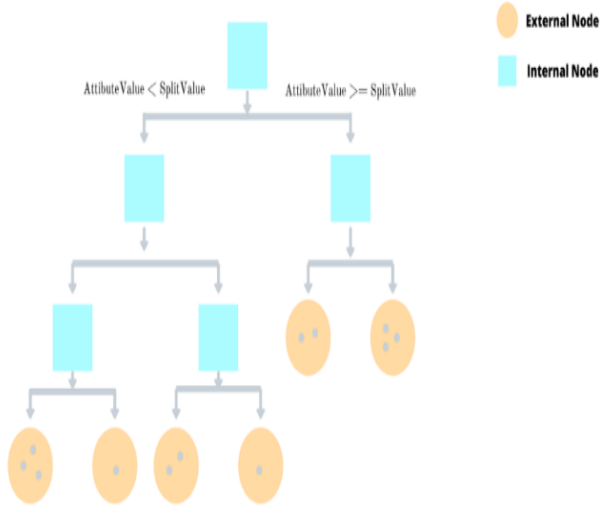

Figure 6: Structure d'un arbre d'isolement avec le point rouge très proche de la racine[32]

Les instances possédant des valeurs de propriété différentes sont séparées au début du processus de partitionnement pour générer des arbres binaires. Ainsi, lorsque ces arbres

binaires produisent ensemble des chemins plus courts dans certaines instances, c'est très probablement une anomalie. En se référant, aux figures 5 et 6, nous observons que le point rouge est facile à séparer par peu de "splits", et en suivant ce processus dans le parcours tracé en rouge, nous ne calculons que deux "splits". En revanche, le point bleu est un peu plus difficile à isoler, il nous faut quatre "split" pour l'isoler comme le montre le graphique de la figure 6 indiqué par le chemin bleu.

Après la construction d'un certain nombre d'arbre binaire, on détermine la longueur moyenne sur ces derniers pour chacune des n instances. Nous concluons que plus la longueur moyenne est courte, plus il y'a une probabilité que l'instance soit une anomalie.

L'algorithme d'arbre d'isolement proposé est le suivant :

Entrée : $\mathcal{E}_{\text{ŧ}}$ échantillon aléatoire de X de taille ŧ

Sortie : arbre binaire

Si $\mathcal{E}_{\text{ŧ}}$ ne saurait être divisé alors :
Retourner un nœud externe de dimension ŧ
sinon
    on suppose Z la liste des attributs caractérisant $\mathcal{E}_{\text{ŧ}}$
    Sélectionner aléatoirement un attribut $q \in Z$
    Sélectionner aléatoirement une valeur de split s_p compris entre le min et le max de l'attribut q
    Xl <= ($\mathcal{E}_{\text{ŧ}}$, $x_q < sp$) : affecter à la branche gauche de l'arbre les instances dont la valeur d'attribut est inférieure à la valeur de séquence s_p.
    Xr <= ($\mathcal{E}_{\text{ŧ}}$, $x_q > sp$) : affecter à la branche droite de l'arbre les instances dont la valeur d'attribut est supérieure à la valeur de séquence s_p.
    Répétition des étapes 4, 5, 6,7 pour isoler davantage les instances des nouveaux nœuds précédentes.
Retourne l'arbre réalisé.

5.4 L'algorithme de la Forêt d'isolement proposé

La forêt d'isolement représente un ensemble d'arbre d'isolement. La construction d'une forêt d'isolement (iforest) se réalise par une combinaison de t arbres binaires.

L'algorithme de construction est le suivant :

Entrée : X, n instances indépendantes, R nombres d'arbre et L taille sous-échantillon
Sortie : Ensemble R arbre binaire

1 Initialisation du Foret d'isolement
2 Pour i allant de 1 à R faire
3 $\mathcal{E}_{ı}$ <= Echantillonner (X, n) : choisir $\mathcal{E}_{ı}$ un échantillon aléatoire de taille $ı$ dans X.
4 Forest <= itree ($\mathcal{E}_{ı}$) : de par l'échantillon $\mathcal{E}_{ı}$, on construit un itree et on l'ajoute à Forest.
4 Fin pour
5 Retourner l'ensemble Foret des R arbres obtenus.

5.5 Calcul du Score d'anomalie

La création de notre forêt d'isolement est insuffisante pour conclure qu'une instance est une anomalie. On doit tout d'abord calculer le score d'anomalie en se basant sur les profondeurs des arbres binaires pour chaque instance.

C'est-à-dire, il suffit de mettre un point dans chacune des arbres binaires du forêt d'isolement et de retracer son itinéraire dans ces arbres jusqu'à ce qu'il atteigne un nœud terminal. Le score ainsi associé à ce point va représenter la moyenne de ce point au niveau des arbres binaires.

Dans une forêt d'isolement, le fait que l'anomalie se positionne la source devient le principal déterminant afin de créer une fonction de notation.

Ainsi le plus long chemin d'un arbre binaire augmente avec n, de même que la longueur moyenne augmente avec log(n).

Pour une instance x le score s est défini par :

$$S(x,n) = 2^{-\frac{E\{(h(x)\}}{c(n)}}$$

Où $E\{(h(x)\}$ est la moyenne des longueurs des chemins parcourus h(x).

c(n) représente un paramètre de normalisation déterminé par :

$$c(n) = 2H\,(n-1) - \{2(n-1)/n\}$$

H(i) est le nombre harmonique estimé par ln(i)+0.5772156649

(Constante d'Euler).

Par conséquent si :

S → 0.5 si $E\{(h(x)\} \to c(n)$

S → 1 si $E\{(h(x)\} \to 0$

S → 0 si $E\{(h(x)\} \to 2^{\frac{n-1}{2\ln(n-1)-2\times 0.52-4\frac{n-1}{n}}} \approx 2^{\frac{n-1}{2\ln(n-1)}}$

Ainsi, nous pouvons observer que le score s est compris entre 0 et 1.

La figure 7 montre la relation existante entre le score s et $E\{(h(x)\}$.

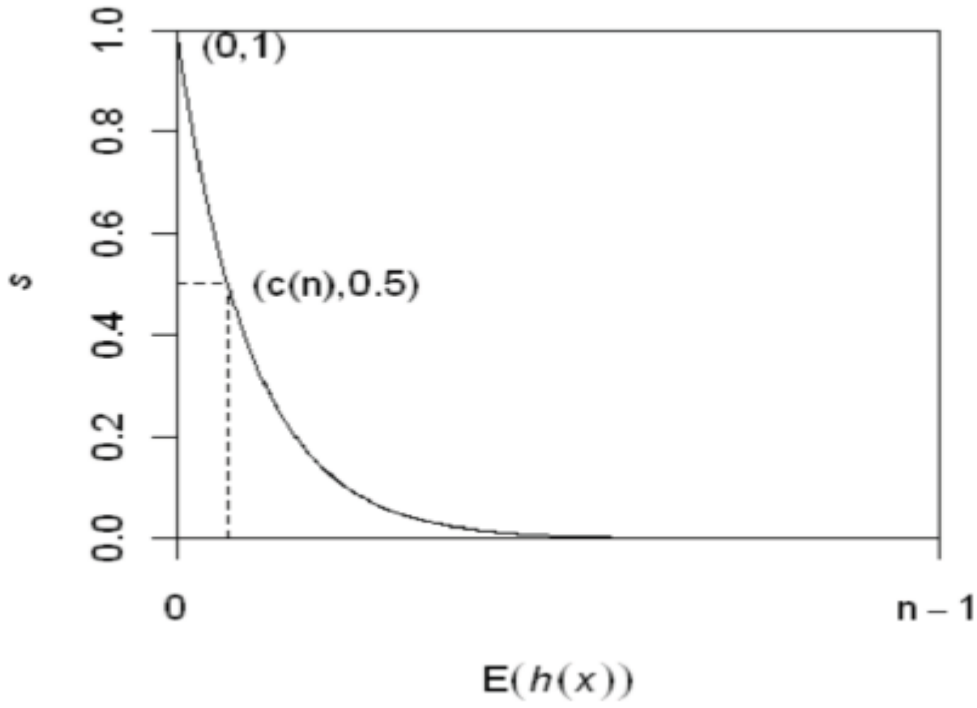

Figure 7: Relation existante entre S et E{(h(x)}

Ainsi, nous pouvons conclure en se basant sur les recommandations de Liu et al. [32]

- Si une instance a un score s inférieur à 0.5, alors il y'a une grande probabilité que cette instance soit normale.
- Si une instance a un score s supérieur à 0.5, alors il y'a une grande probabilité que cette instance soit une anomalie.

Conclusion

Dans cette section, nous avons présenté le fonctionnement de notre modèle ainsi que les différents algorithmes pour la création des différents arbres et forêts d'isolement.

Afin de prouver l'efficacité de l'approche proposé, nous allons présenter l'application de notre modèle au le prochain chapitre en utilisant la base de données V2G développée au sein de notre laboratoire.

Chapitre 6 – Expériences et Résultats

6.1 Introduction

Ce chapitre décrit la méthodologie utilisée, il comprend une description des outils utilisés pour mettre en œuvre l'IDS ainsi que les méthodes de test et d'évaluation des performances.

6.2 Recherche générale

Après une étude d'état de l'art sur les détections d'intrusions dans les réseaux VANET et V2G, nous allons proposer une solution de détection d'anomalie basée sur la forêt d'isolement.

En accédant aux spécificités de la norme ISO 15118, nous nous sommes efforcés de comprendre comment fonctionnent les différents acteurs d'une session de communication V2G, quels étaient les différents types de messages et quelles fonctionnalités ils fournissent. Ainsi, nous avons pu obtenir une vue plus approfondie pour proposer un système de détection d'intrusions

6.3 Environnements Logiciels

6.3.1 Présentation du simulateur MiniV2G

MiniV2G [34] est un émulateur open source pour simuler la recharge de véhicules électriques (EVC) construit sur Mininet et RiseV2G. MiniV2G peut reproduire avec haute fidélité une architecture V2G pour simuler facilement un processus de charge de VE. MiniV2G peut être utilisé pour étudier la communication V2G et développer des attaques et des contre-mesures applicables à des systèmes réels.

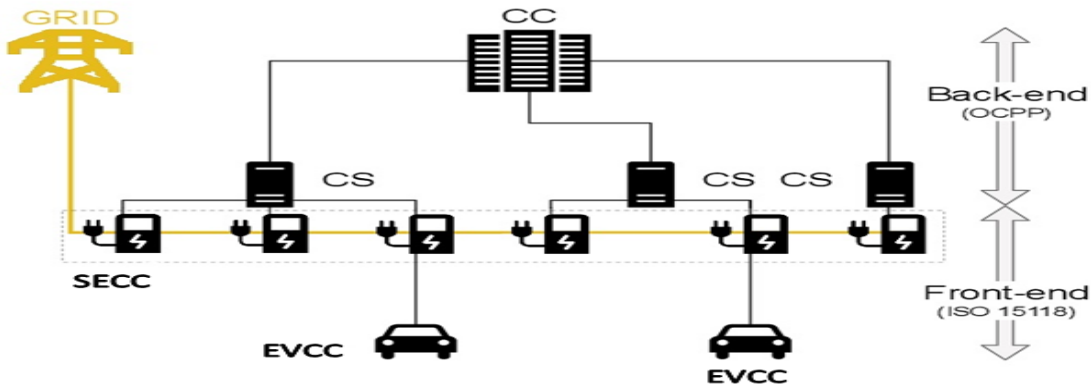

Figure 8: Modélisation d'une communication V2G avec MiniV2G

6.3.2 Inspection des paquets à l'aide de Wireshark

Wireshark est un logiciel gratuit permettant de prendre des captures instantanées du trafic réseau. Il peut être utilisé pour renifler, inspecter, filtrer et analyser les protocoles réseau en temps réel et prend en charge le filtrage hors ligne des captures de paquets. Lors du développement d'applications réseau telles que les IDS, Wireshark est nécessaire pour inspecter les entrées et les sorties des applications, en plus de sa capacité de capturer les scénarios de trafic générés. Wireshark est particulièrement important pour nous car il nous permet d'inspecter les paquets V2G envoyés entre le serveur SECC et le client EVCC. Cela permet de déterminer l'ordre, la taille et la fréquence des messages.

6.3.3 Libpcap

Une autre bibliothèque C importante utilisée dans le projet est libpcap (le nom complet du package est libpcap0.8-dev). Son API fournit des appels de fonction pour la lecture de fichiers pcap ainsi que la capture en direct de paquets avec des capacités de filtrage.

6.3.4 Boite Virtuelle

Pour le développement, nous avons utilisé VMWare pour exécuter le simulateur MiniV2G.

6.3.5 Google Colab

Colaboratory, souvent appelé « Colab », est un produit de Google Research. Colab permet à quiconque d'écrire et d'exécuter le code Python de son choix via un navigateur. Il s'agit d'un

environnement particulièrement adapté à l'apprentissage automatique, à l'analyse de données et à l'éducation.

6.4 Méthode d'obtention de la base de données V2G

6.4.1 Session V2G

Pour établir une session V2G, nous avons utilisé le logiciel client et serveur sur les exemples de test disponibles dans le code source MiniV2G.Les exemples accessibles dans MiniV2G utilisent des flux d'octets exclusivement pour simuler une session entre le client et le serveur, il n'y a pas de différenciation entre le serveur et le client. Comme nous voulions simuler des communications réelles conformément à la spécification ISO 15118, les appels spécifiques au serveur et au client ont été divisés en fichiers séparés (evcc_client et secc_server). En fonction de l'entrée de ligne de commande au démarrage du fichier principal exécutable ; un EVCC ou un SECC est démarré. Pour exécuter une session V2G, on démarre d'abord le SECC, puis l'EVCC, tel qu'il est activé en tant que processus distincts. Avec l'intégration supplémentaire d'UDP et TCP en utilisant IPv6, nous avons pu exécuter avec succès des sessions V2G.

6.4.2 Client EVCC

Le client simule un EVCC et a été implémenté en tant que partie initiatrice lors d'une session V2G conformément à la spécification ISO 15118. Le client est activé via une fenêtre de terminal, il prend les paramètres suivants : interface réseau, adresse IPv6 du serveur, mode de transfert d'énergie (AC ou DC). Il existe deux exemples de sessions statiques différentes qui peuvent être réalisées, en utilisant la charge AC ou DC.

6.4.3 Serveur SECC

Le serveur SECC est implémenté comme un auditeur passif avec une gestion des erreurs limitées et des capacités de prise de décision adaptées à V2G. Quel que soit le contexte de ce message dans la communication V2G, il répond toujours à la demande du client par une réponse, par exemple, si une demande SessionStop est reçue, le serveur peut mettre fin à la session en répondant par une réponse SessionStop, qu'il ait mis fin à la séance ou non. Dans

une implémentation réelle de la spécification ISO 15118, le serveur ne se comporterait pas de cette façon. Le serveur fonctionne de la même manière que le client.

6.4.4 Intégration des attaques

Nous devons intégrer des scénarios d'attaque dans notre implémentation pour pouvoir tester la sécurité d'IDS. Considérons ces exemples : nous voulons effectuer une attaque en supprimant les réponses du serveur, refusant effectivement le service au client, nous voulons submerger le serveur avec beaucoup de confusion chez le client. Compte tenu de ces exemples, il est clair que les deux parties doivent s'écarter du comportement de session normal. En d'autres termes, nous intégrons la fonctionnalité d'attaque directement dans le serveur et le client, respectivement. L'attaque provient entièrement du client EVCC (par exemple, l'envoi de plusieurs requêtes au serveur), et le client déclenche également des exploits côté serveur pour certains scénarios d'attaque (par exemple, lorsqu'il répond tardivement à un message).

Nous avons aussi intégré l'attaque de MITM, où il y'a un client lamda qui intercepte les informations transmises entre le serveur SECC et le client EVCC.

Après avoir sniffé la communication client-serveur de la communication V2G grâce à l'outil wireshark, nous avons obtenu notre base de données au format pcap et transformé en csv grâce à l'outil CiCflowmeter.

6.5 Implémentation de la méthode du Forêt d'isolement

Lors de la génération de la base de données V2G, une colonne « ATTAQUE » a été ajoutée.

- Pour les données relatives aux scénarios sans attaque, la variable ATT est égale à 0.
- Pour les données relatives aux scénarios avec attaques (MITM, DOS), la variable ATT est égale à 1.

6.5.1 Sélection de deux attributs plus corrélés avec l'attribut ATT

Afin de réaliser notre projection, on doit sélectionner deux attributs plus significatifs, autrement dit deux variables qui seront plus affectées quand il y'a une attaque.

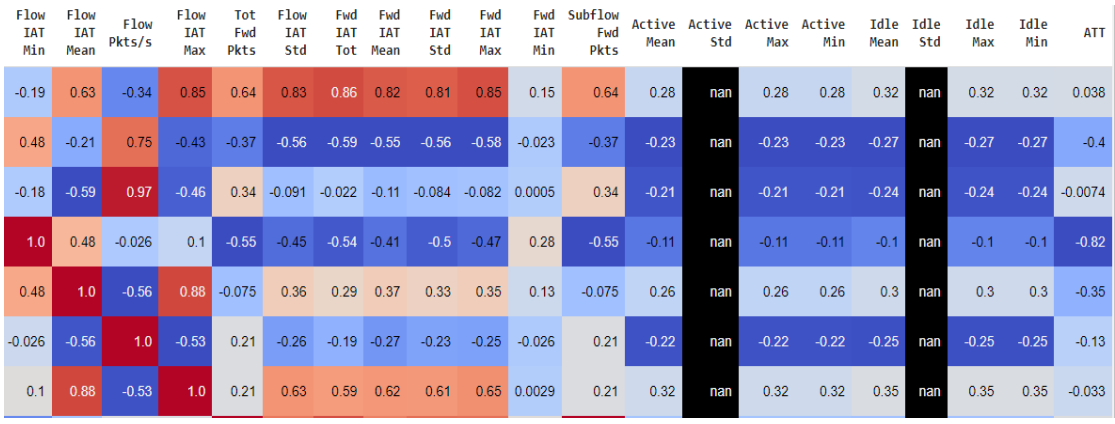

Figure 9: Corrélation entre les différents attributs.

Nom de Fonction	Description
Flow IAT min	Temps minimum entre deux flux
Flow IAT mean	Temps moyen entre deux paquets envoyés
Flow pkts/s	Débit des paquets qui est le nombre de paquets transférés par seconde
Flow IAT Max	Temps maximum entre deux flux
Tot Fwd pkts	Total des paquets dans le sens direct
Flow IAT Std	Temps d'écart type deux flux
Fwd IAT Tot	Temps total entre deux paquets envoyés dans le sens direct
Fwd IAT Mean	Temps moyen entre deux paquets envoyés dans le sens direct
Fwd IAT Std	Temps d'écart type entre deux paquets envoyés dans le sens direct
Fwd IAT Max	Temps maximum entre deux paquets envoyés dans le sens direct
Fwd IAT Min	Temps minimum entre deux paquets envoyés dans le sens direct
Subflow Fwd pkts	Le nombre moyen de paquets dans un sous-flux dans le sens direct
Active Mean	Temps moyen où un flux était actif avant de devenir inactif
Active Std	Écart type temps où un flux était actif avant de devenir inactif
Active Max	Durée maximale où un flux était actif avant de devenir inactif
Active Min	Temps minimum où un flux était actif avant de devenir inactif
Idle Mean	Temps moyen où un flux était inactif avant de devenir actif
Idle Std	Écart type temps où un flux était inactif avant de devenir actif
Idle Max	Temps maximum où un flux était inactif avant de devenir actif
Idle Min	Temps minimum où un flux était inactif avant de devenir actif
ATT	Attaque Dos et Homme du milieu

Tableau 3: Liste des fonctionnalités du trafic

On observe de par nos différentes variables que l'attribut **Fwd_iat_tot** (temps total entre deux paquets envoyés dans le sens direct) et **Fwd_iat_std**(temps d'écart type entre deux paquets envoyés dans le sens direct) sont plus corrélées par rapport à l'attribut ATT avec un score de 0.21.

6.5.2 Projection des attributs

- **Visualiser la distribution gaussienne bivariée**

La distribution gaussienne (ou distribution normale) étant l'une des distributions de probabilité les plus essentielles. Elle nous permet de visualiser et de comprendre à l'aide de tracés géométriques appropriés.

La distribution bivariée peut aussi étendre n'importe quel nombre de dimension appris par cette dernière.

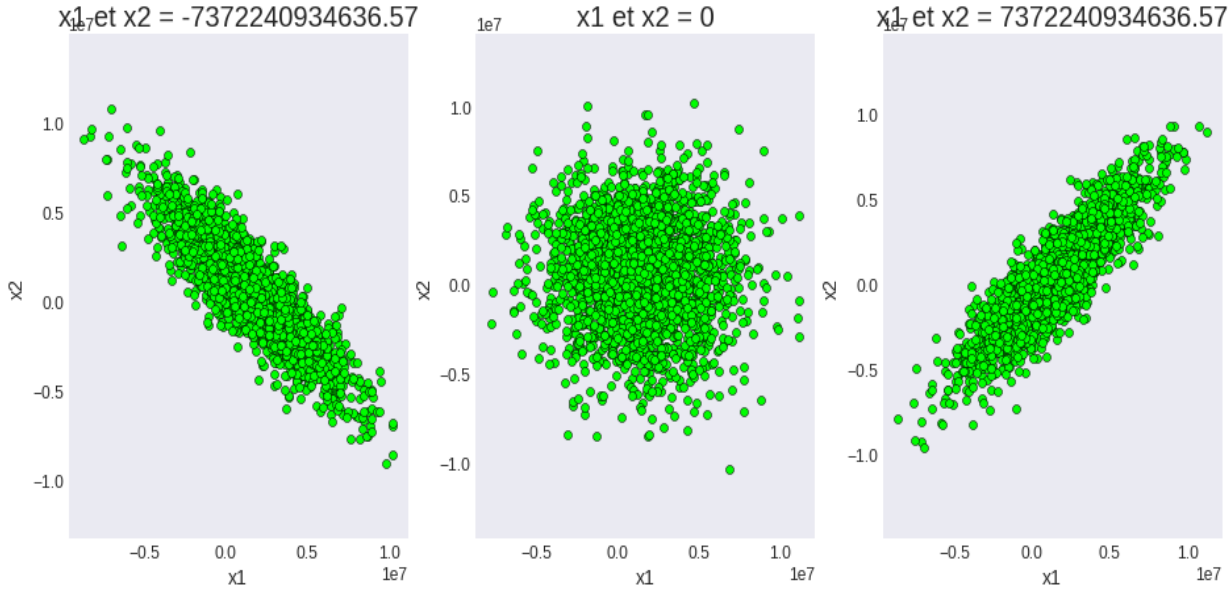

Figure 10: Représentation Gaussienne des différentes distributions bivariées

Dans ce cas de figure, nous avons choisi quand la covariance

x1 et x2 = 0 vu la forte densité des données.

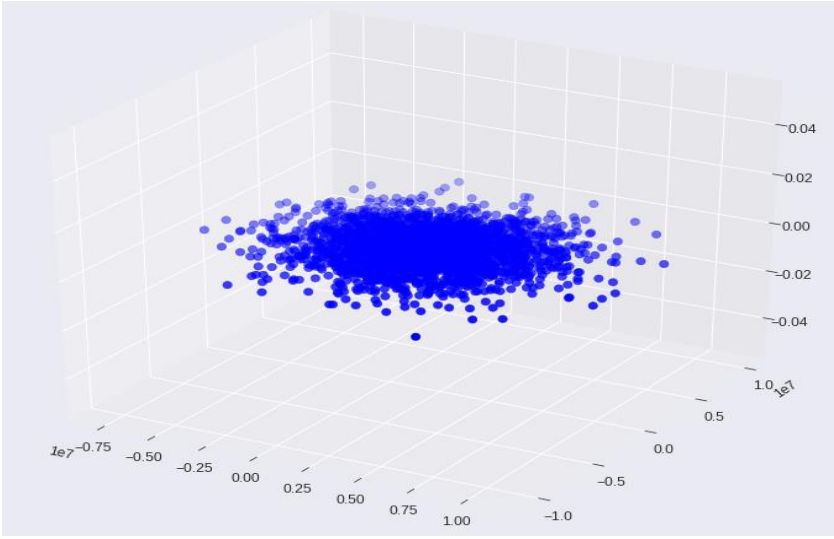

Figure 11: Représentation en 2D des attributs Fwd_iat_tot et Fwd_iat_std

6.5.3 La construction de l'arbre de d'isolement

Étant donné que les anomalies peuvent être isolées et ont tendance à être plus proches de la racine de l'arbre de décision, nous avons construit l'arbre de décision sans autres points de division jusqu'à ce qu'il atteigne une certaine hauteur. Cette hauteur est la colonne dont nous sommes (presque) sûrs qu'il n'y aura pas d'anomalies.

```
{'Attribut1 <= -601722.2729171384': [{'Attribut2 <= 7319668.3670821525': [-8563343.013175428,
                                                                          7369803.825204297]},
                                    {'Attribut1 <= 4220497.769130932': [-8525338.95739876,
                                                                        -10418978.242094412]}]}
```

Figure 12: Représentation de l'arbre d'isolement

À la figure 12, nous avons la structure de notre arbre d'isolement avec les deux attributs sélectionnés en haut pour la projection des splits dans l'échantillon. La profondeur maximale pour ce cas étant 2, afin de mieux observer la structure. Mais la capacité de détection est en phase avec la profondeur de l'arbre. Plus la profondeur est grande, plus la capacité de détection est importante.

6.5.4 Construction de la forêt d'isolement

Après avoir obtenu notre arbre nous devons en faire une forêt.

IForest divise les données en hyper rectangles construits aléatoirement à un temps de complexité linéaire, c'est-à-dire le temps d'exécution de l'algorithme est une fonction linéaire de taille de données n utilisée à l'entrée.

L'algorithme de la forêt d'isolement a des hyper paramètres, une taille de sous-échantillon **v** et le nombre d'arbres binaires t.

Nous avons constaté qu'à mesure que la taille de l'échantillon augmente, iForest détecte les anomalies de manière fiable.

Dans une étude empirique, fixer le paramètre de l'échantillon à $2^8 = 258$ est généralement suffisant pour détecter des anomalies dans de grands ensembles de données.

Concernant le nombre d'arbres binaires, l'algorithme fonctionne bien avec la valeur de n_trees = 100, qu'on utilise comme valeur par défaut lors de l'exécution de l'algorithme.

Une portion de iForet produit est donnée ci-dessous :

```
1 IsolationForest(X,n_trees=20,max_depth=2)
```

```
{'Attribut2 <= -2596669.8209175216': [{'Attribut2 <= -6477583.411968709': [-8525338.95739876,
    -5983345.242097571]},
  {'Attribut2 <= 4416174.330186017': [-2562945.334569027,
    4464256.453287604]}]},
{'Attribut2 <= -487204.41336383857': [{'Attribut2 <= -3641832.1229578247': [-8525338.95739876,
    -3626402.3534892504]},
  {'Attribut2 <= 2142494.5003202446': [-477712.1216341815,
    2154502.6466298685]}]},
{'Attribut1 <= -5264372.398631588': [{'Attribut2 <= 497289.32683109445': [-3928724.3094457197,
    1176559.703953546]},
  {'Attribut2 <= 2223354.07036913': [-8525338.95739876, 2229806.252222343]}]},
{'Attribut1 <= 1540418.974530547': [{'Attribut1 <= -132877.57274302002': [-4052313.888643132,
    -5759923.603313977]},
  {'Attribut1 <= 5476216.634993935': [-8525338.95739876,
    -5932235.893215116]}]},
{'Attribut1 <= 162841.22342635877': [{'Attribut1 <= -2794357.2038493706': [-3928724.3094457197,
    -4351262.639312472]},
  {'Attribut1 <= 3998382.259533831': [-8525338.95739876,
    -5932235.893215116]}]},
{'Attribut1 <= 844946.7940743221': [{'Attribut1 <= -2320509.8629902545': [-4052313.888643132,
    -5759923.603313977]},
  {'Attribut1 <= 954428.3683163209': [-5297997.441856726,
    -8525338.95739876]}]},
```

Figure 13: Structure de la Forêt d'isolement

Dans notre cas, nous choisissons le paramètre n_trees=20 pour un meilleur visionnement, où n_trees représente le nombre d'arbre qui constitue la forêt d'isolement et max_depht la profondeur maximale de la forêt.

Nous comptons ensuite le nombre de nœuds que chaque instance traverse avant d'être isolée ou d'atteindre la profondeur maximale. Ensuite, on détermine la moyenne des longueurs des chemins sur l'ensemble des n instances de l'arbre binaire. Plus la longueur moyenne est courte, plus cette instance est probablement une anomalie.

La figure 14 ci-dessous montre la mise en évidence, d'une instance jugée anormale dans l'échantillon :

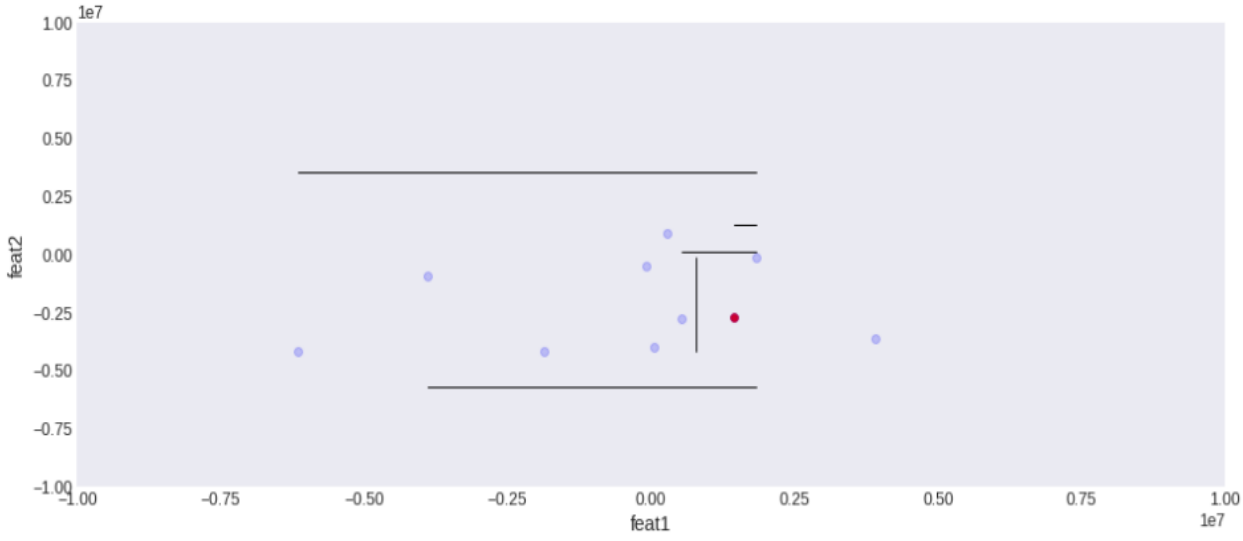

Figure 14: Mise en évidence d'une instance anormale

On détermine ensuite la proportion d'instance normale et d'anomalie présente dans notre échantillon.

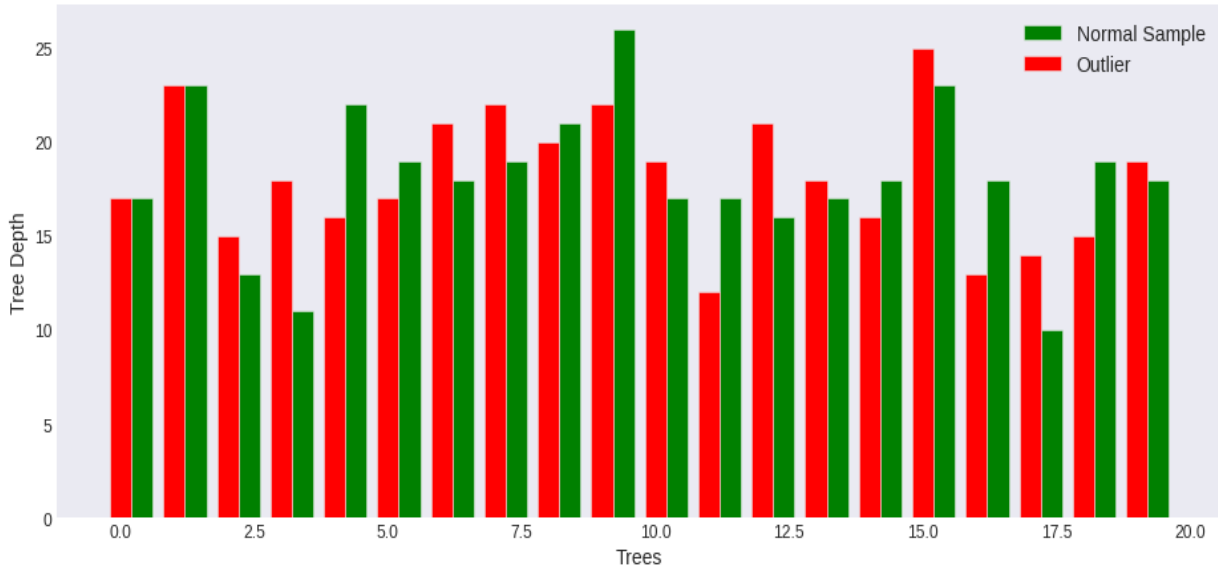

Figure 15: Instance normale et anomalie dans un échantillon

À la figure 15 ci-dessus, nous avons la représentation des instances normales (vert) et anomalie (rouge) sur chaque arbre produit. L'ensemble constitue notre forêt avec 20 arbres.

Ensuite, nous devons déterminer le score d'anomalie pour appuyer s'il y'a présence ou non d'anomalie dans l'échantillon de donnée en se basant sur la formule évoquée au chapitre 5.

Ainsi, nous obtenons un score d''anomalie de 97%, en effet ce score est très proche de 1, d'où on peut conclure que notre échantillon possède des anomalies.

6.6 Evaluation et Comparaison avec d'autres algorithmes de détection d'anomalie

L'IDS peut être considéré comme un système de classification binaire dans lequel les paquets sont classés comme positifs (détection) ou négatifs. Une classification positive signifie qu'une attaque est en cours dans le système, tandis qu'une classification négative signifie tous les autres événements possibles (y compris le trafic normal et les messages d'erreur). La matrice de confusion illustrée à la figure 7 illustre les quatre résultats du système de classification binaire et peut être utilisée pour calculer les paramètres de performances. Les quatre résultats sont vrai négatif, faux négatif, vrai positif et faux positif.

		Predicted Class	
		Normal	Attack
Actual Class	Normal	True Negative (TN)	False Positive (FP)
	Attack	False Negative (FN)	True Positive (TP)

Figure 16: Matrice de confusion pour classification binaire

Taux négatif et Positifs

- Le taux de vrai positif (TPR) ou rappel est défini comme les attaques correctement appliquées contre le système. Par conséquent, un classificateur qui ne produit aucun faux négatif atteindra un rappel de 1,0.

$$TPR\,(Recall) = \frac{TP}{TP + FN}$$

- Le taux de faux positifs (FPR) définit les attaques incorrectement diffusées contre le système ; les paquets normaux étant classés comme des attaques. Le FPR optimal serait un taux de 0 sans faux positif.

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

- Le taux de faux négatifs (FNR) représente le taux auquel une attaque est faussement refusée au système ; l'attaque est identifiée comme un paquet normal. Semblable au FPR, le meilleur cas de performance FNR a un rapport de 0 et aucun faux positif.

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

Précision

La précision est définie comme la proportion d'identifications positives (attaques) qui ont été correctement détectées. Un score de précision parfait est de 1,0 sans faux positifs.

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

F-Score

Le score F est une mesure de la précision d'un test, de plus c'est la moyenne harmonique de précision et de rappel.

$$F_\beta = (1 + \beta^2)\frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}$$

Pour l'évaluation de notre modèle, nous utilisons un support égal à 97 avec comme nombre d'erreur égal à 17, un taux de vrai positif 39 et de vrai négatif 24 comme illustré à la figure 17 ci-dessous :

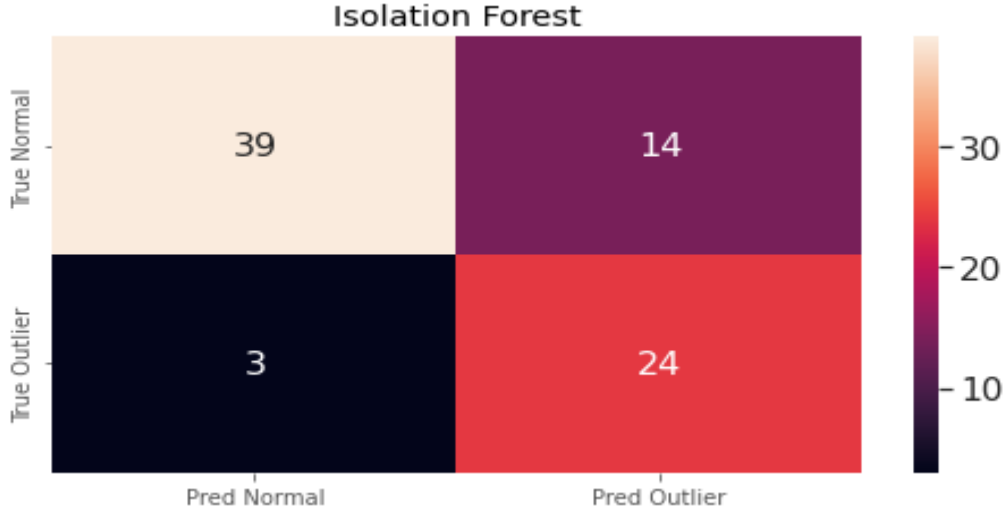

Figure 17: Evaluation du modèle de forêt d'isolement

Après avoir évalué les scores métriques de l'algorithme, nous obtenons les résultats classés dans le tableau suivant :

AUC score	Précision	recall	F1-score	Durée
81%	93%	89%	82%	0.37 s

Tableau 4: Evaluation des performances du modèle de Forêt d'isolement

Ces résultats obtenus montrent que l'algorithme est efficace en termes de détection d'instance normale comme d'anomalie avec des données V2G. En effet, l'algorithme de la forêt d'isolement étant plus distingué sur sa capacité à détecter sur des données comportant de grands supports, ainsi plus le support de donnée devient grand plus sa capacité de détection devient importante.

D'autres algorithmes de détection ont été testé avec l'ensemble de données V2G afin de faire la comparaison en termes de performance.

❖ Local Outlier Factor

Local Outlier Factor (LOF) est un algorithme de détection d'anomalies utilisant une méthode non supervisée. Sa particularité étant de calculer l'écart d'une densité d'une instance vis à vis de ses voisins. En effet, il considère comme anomalie les échantillons ayant une densité inférieure à leurs voisins.

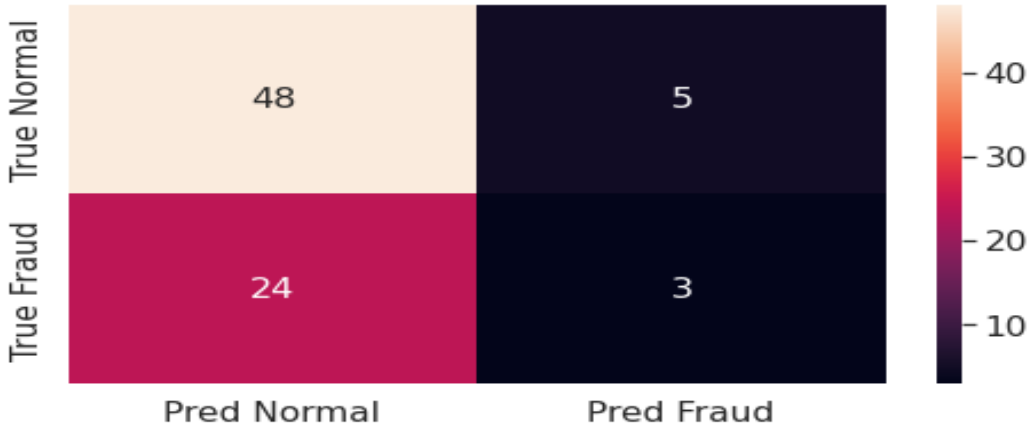

Figure 18: Evaluation du modèle Local Outlier Factor

AUC score	Précision	recall	F1-score	Durée
50%	67%	91%	77%	0.112s

Tableau 5: Evaluation des performances du modèle de Local Outlier Factor

De par ces résultats l'algorithme Local outlier factor possède un taux de nombre d'erreur égal à 29 sur un support de 97.

Ainsi ces performances sont moins élevées par rapport à la méthode de l'isolation forêt sur la base de données V2G, mais néanmoins il possède une durée d'exécution moins grande que celle de la forêt d'isolement ; et cela s'explique du fait qu'à l'application de la forêt d'isolement, plus le nombre d'arbre utilisé dans une forêt augmente, plus la durée d'exécution de l'arbre devient importante.

- **Fast-MCD (Minimum Covariance Determinant)**

Fast-MCD est un algorithme très utile pour détecter des valeurs aberrantes dans une base de données. Sa méthode se base sur une supposition que les instances normales sont obtenues sur la base d'une distribution gaussienne. Il suppose de même que la contamination de la base de données provient des données aberrantes non issues de cette distribution.

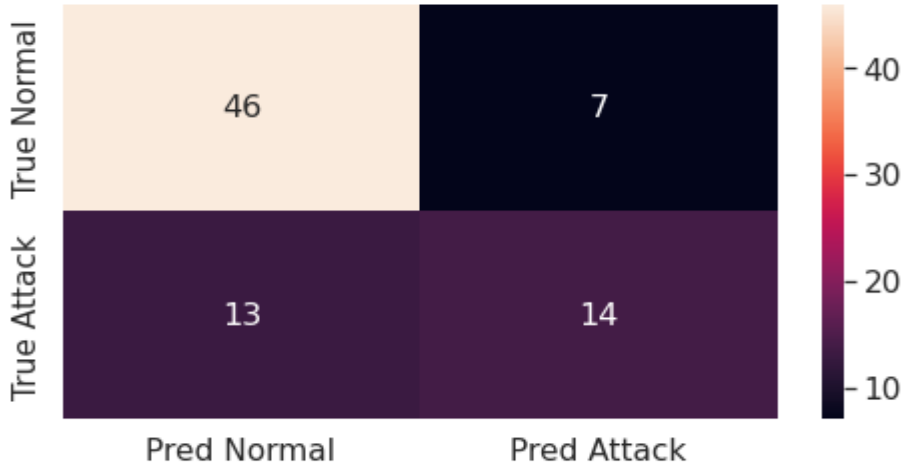

Figure 19: Evaluation du modèle fast-MCD

AUC score	Précision	recall	F1-score	durée
69%	78%	87%	82%	0.129s

Tableau 6: Evaluation des performances du modèle Fast-MCD

Suite au test sur l'algorithme fast-MCD sur la base de données V2G, ce dernier comporte 20 erreurs sur un support de 97. En effet, ses performances en termes de détection sont moins importantes que celles de l'algorithme de la forêt d'isolement sur un ensemble de données V2G comme en illustre le tableau 6 ci-dessus.

	AUC score	**Precision**	**Recall**	**F1-score**	**Duration**
Isolation Forest	81%	93%	89%	82%	0.37 s
Local Outlier Factor	50%	67%	91%	77%	0.112s
fast-MCD	69%	78%	87%	82%	0.129s

Tableau 7: Tableau récapitulatif des différents résultats.

6.7 Conclusion

Dans ce chapitre, nous avons effectué la mise en expérience de notre modèle sur la base de données V2G en appliquant les différents algorithmes proposés. Nous avons aussi testé d'autres algorithmes d'apprentissage non supervisé sur le même jeu de donnée afin de comparer les performances.

Nous pouvons déduire suite aux expériences et comparaisons que la détection d'anomalie utilisant la méthode de la forêt d'isolement possède des performances plus importantes, avec une faible complexité de calcul et une grande applicabilité aux données complexes. Dans nos travaux futurs, nous allons tenter d'améliorer nos propositions afin d'obtenir des performances plus importantes, mais aussi réaliser des évaluations sur des scénarios d'autres attaques.

Conclusion Générale et Perspectives

Dans cette thèse de maîtrise, nous avons essayé de proposer des fonctions précieuses pour un système de détection d'intrusion (IDS) dans un environnement V2G. Dans cette dynamique, nous avons étudié les spécifications de charge V2G pour établir les exigences et les fonctionnalités de mise en œuvre d'un IDS et l'intégration d'une inspection approfondie des paquets. Nous avons proposé ensuite de mettre en œuvre une preuve de concept pour un IDS V2G avec l'utilisation de méthodes basées sur les anomalies.

La détection basée sur les anomalies se concentre sur les écarts par rapport au comportement spécifié dans une session, qui sont considérés comme anormaux par rapport à la spécification définissant une session V2G : ISO 15118.

Dans notre approche, nous avons collecté des données à partir de différents exemples de sessions V2G et sur la base de ces données, nous avons classé le paquet comme une menace ou un paquet normal. Notre moteur de détection d'anomalie basé sur les forêts d'isolement évalue la fréquence, la durée de charge utile et le délai d'expiration attendus pour la demande-réponse de chaque type de message. De plus, la détection d'un paquet dépend de deux seuils différents : la tolérance et la limite inférieure/supérieure. Les deux seuils affectent le nombre de paquets autorisés à ne pas être détectés.

Pour évaluer notre implémentation, nous avons construit des tests avec différents scénarios d'attaques. Les résultats montrent que notre méthode de détection d'anomalie basée sur les forêts d'isolement peut détecter avec succès les attaques et constitue une solution prometteuse pour détecter les menaces dans un environnement de charge V2G.

Lors de l'évaluation de notre IDS, nous avons constaté que la méthode de détection d'intrusion basée sur les anomalies se limite sur les comportements atypiques, mais en le combinant avec une méthode fournissant les mêmes spécifications d'un détecteur d'intrusion basé sur les signatures, elles peuvent fonctionner en tant que méthode hybride pour détecter les menaces connues et invisibles.

Nos tests montrent que lors des sessions V2G normales avec peu de faux positifs, un seuil de tolérance accru augmentera la précision, tout en diminuant légèrement le taux de vrais positifs. Il convient d'approfondir l'étude et l'évaluation des deux seuils afin de déterminer leur importance pour la réalisation de la détection fondée sur les anomalies.

Une prochaine étape serait d'implémenter l'IDS dans un environnement réel grandeur nature et d'évaluer ses performances sur des cas d'utilisation réels ainsi que de tester des scénarios d'autres attaques.

Références bibliographiques

[1].Cenex. Recharge de VÉHICULES ÉLECTRIQUES url: https:// www.cleantech.com/ev-charging-software-and-grid-services/.

[2]. https://www.iso.org/fr/standard/55365.html.

[3]. Département des nouvelles de PG&E. Pacific Gas and Electric Company Ener- gizes Silicon Valley avec la technologie Vehicle-to-Grid. dans: PG&E News Department 415.4. 2007.

[4].ISO 15118:2014(E). Véhicules routiers - Interface de communication véhicule-réseau - Partie 2: Exigences relatives aux protocoles de réseau et d'application (ISO 15118-2:2014). Norme 32000– 1:2008. Genève, Suisse: Organisation internationale de normalisation, 2008.

[5].W3C. Format EXI (Efficient XML Interchange) 1.0 (deuxième édition). [En ligne; consulté le 6 janvier 2020]. 2014. url: https://www.w3.org/ TR/exi/.

[6]. Miller C et Valasek C. Exploitation à distance d'un véhicule pas-senger inchangé. [En ligne; consulté le 2 octobre 2019]. Août 2015.url:http://illmatics.com/Remote%5C%20Car%5C%20Hacking.pdf

[7]. https://info-ceh.blogspot.com/2011/07/intrusion-detection-system-ids.html.

[8]. Liu, Fei Tony ; Ting, Kai Ming; Zhou, Zhi-Hua (décembre 2008). "Forêt d'isolement". 2008 Huitième Conférence internationale IEEE sur l'exploration de données : 413–422.

[9]. Attanasio, Luca and Conti, Mauro and Donadel, Denis and Turrin, Federico, "MiniV2G: an electric vehicle charging emulator", Proceedings of the 7th ACM workshop on the security of cyber-physical systems, 65-73.

[10]. Aldwairi M, Abu-Dalo AM, and Jarrah M. Pattern matching of signaturebased IDS using Myers algorithm under MapReduce framework. In: EURASIP Journal on Information Security. 2017, pp. 1–11.

[11]. Mazen Kharbutli, Monther Aldwairi, and Abdullah Mughrabi. Function and Data Parallelization of Wu-Manber Pattern Matching for Intrusion Detection Systems. In: Network Protocols and Algorithms 4. Sept. 2012, pp. 46–61.

[12]. X. Clotet, J. Moyano et G. León, "Un IDS basé sur les anomalies en temps réel pour la détection des cyberattaques au niveau des processus industriels des infrastructures critiques", Crit. Infrastructure. Protection, vol. 23, p. 11-20, décembre 2018.

[13]. W. Aoudi, M. Iturbe et M. Almgren, "La vérité sortira : détection au niveau du processus basée sur les départs des attaques furtives contre les systèmes de contrôle", ACM Conf. Calcul. Commun. Sécurisé , p. 817-831, 2018.

[14]. D. Shalyga, P. Filonov et A. Lavrentyev, "Détection d'anomalies pour un système de traitement de l'eau basé sur un réseau neuronal avec optimisation automatique de l'architecture", arXiv:1807.07282 , 2018, http://arxiv.org/ abs/1807.07282.

[15]. M. Kravchik et A. Shabtai, "Détection des cyberattaques dans les systèmes de contrôle industriels à l'aide de réseaux neuronaux convolutifs", Proc. Atelier Cyber-Physical Syst. Sécurisé Confidentialité (CPS-CPS), p. 72-83, 2018.

[16]. J. Inoue, Y. Yamagata, Y. Chen, CM Poskitt et J. Sun, "Détection d'anomalies pour un système de traitement de l'eau par apprentissage automatique non supervisé", Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW) , pp. 1058-1065, novembre 2017.

[17]. D. Li, D. Chen, B. Jin, L. Shi, J. Goh et S.-K. Ng, "MAD-GAN : Détection d'anomalies multivariées pour les données de séries chronologiques avec des réseaux antagonistes génératifs" dans Réseaux de neurones artificiels et apprentissage automatique, Cham, Suisse : Springer, pp. 703-716, 2019.

[18]. Z. Ye, A. Gilman, Q. Peng, K. Levick, P. Cosman and L. Milstein, "Comparison of neural network architectures for spectrum sensing", arXiv:1907.07321, 2019, [online] Available: http://arxiv.org/abs/1907.07321.

[19]. Z. Yang, S. Yu, W. Lou, and C. Liu. p2: Privacy-preserving communication and precise reward architecture for v2g networks in smart grid. IEEE Transactions on Smart Grid, 2(4):697–706, 2011.

[20]. H. Li, G. D´an, and K. Nahrstedt. Lynx: Authenticated anonymous realtime reporting of electric vehicle information. In Smart Grid Communications (SmartGridComm), 2015 IEEE International Conference on, pages 599–604. IEEE, 2015.

[21]. Vaidya B. , Makrakis D. , Mouftah HT : ' Mécanisme de sécurité pour les infrastructures multi-domaines vehicle-to-grid '. Proc. IEEE GLOBECOM, 2010 , p. 1 à 5.

[22]. Liu H. , Ning H. , Zhang Y. et al. : « Préservation de la confidentialité en fonction du rôle pour les réseaux V2G sécurisés dans le réseau intelligent », IEEE Trans. Inf. Sec. , 2014 , 9 , (2), p. 208 – 220

[23]. Wan Z. , Zhu W. , Wang G. : « PRAC : protection efficace de la vie privée pour les communications véhicule-réseau dans le réseau intelligent », Comput. Sécurisé , 2016 , 62 , p. 246 – 256.

[24]. B. Vaidya, D. Makrakis, and H. T. Mouftah. Efficient authentication mechanism for pev charging infrastructure. In Communications (ICC), 2011 IEEE International Conference on, pages 1–5. IEEE, 2011.

[25]. R. Kolandaisamy, RM Noor, I. Ahmedy et al., "Une approche d'analyse de flux multivariante pour détecter et atténuer les attaques DDoS dans les réseaux ad hoc véhiculaires," Wireless Communications and Mobile Computing , vol. 2018, Article ID 2874509, 13 pages, 2018.

[26]. M. Raya, P. Papadimitratos, I. Aad, D. Jungels, and J.-P. Hubaux, “Eviction of Misbehaving and Faulty Nodes in Vehicular Networks,” IEEE Journal on Selected Areas in Communications, 2007.

[27]. Hasrouny, Hamssa, et al. "VANET security challenges and solutions: A survey." Vehicular Communications 7 (2017): 7-20.

[28]. Al-Terri, Doaa, et al. "Cooperative based tit-for-tat strategies to retaliate against greedy behavior in VANETs." Computer Communications 104 (2017): 108-118.

[29]. Yi Zeng et al, ‘DeepVCM: A Deep Learning Based Intrusion Detection Method in VANET’ 2019 IEEE 5th Intl Conference on High Performance and Smart computing.

[30]. Safi, Qamas Gul Khan, et al. "PIaaS: Cloud-oriented secure and privacyconscious parking information as a service using VANETs." Computer Networks 124 (2017): 33-45.

[31]. Zaidi, Kamran, et al. "Host-based intrusion detection for vanets: a statistical approach to rogue node detection." IEEE transactions on vehicular technology 65.8 (2016): 6703-6714.

[32]. F. T. Liu, K. M. Ting and Z.-H. Zhou, ″On detecting clustered anomalies using SCiForest″, Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases, pp. 274-290, 2010.

[33] Attanasio, Luca et Conti, Mauro et Donadel, Denis et Turrin, Federico, "MiniV2G : un émulateur de charge de véhicule électrique", Actes du 7e atelier ACM sur la sécurité des systèmes cyber-physiques, 65-73, 2021.