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

DIEUDONNE MIZONZA BANTIKO

OPTIMISATION DU SYSTÈME DE DÉTECTION D'INTRUSIONS (IDS) DANS LE RÉSEAU VÉHICULAIRE V2G À L'AIDE DES RÈGLES D'ASSOCIATION MAXIMALES ET DE LA RÉGRESSION LOGISTIQUE

Août 2022

PRÉSENTATION DU JURY

Ce mémoire a été évalué par un jury composé de :

Boucif Amar Bensaber, directeur de recherche.
Professeur au département de Mathématiques et d'Informatique,
Université du Québec à Trois-Rivières.

Ismail Biskri, évaluateur.
Professeur au département de Mathématiques et d'Informatique,
Université du Québec à Trois-Rivières.

Mhamed Mesfioui, évaluateur
Professeur au département de Mathématiques et d'Informatique,
Université du Québec à Trois-Rivières.

REMERCIEMENTS

La réalisation de ce projet a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma gratitude.

Je remercie Dieu le tout-puissant pour ses hauts faits.

Mes sincères gratitudes en guise de reconnaissance au Professeur Boucif Amar Bensaber, Directeur de ce mémoire à qui j'adresse un merci particulier pour sa disponibilité et surtout ses stratégies, méthodes et rigueur utilisées pour m'amener à réaliser ce travail qui est le fruit de nos efforts. Qu'il trouve dans ce travail ma profonde considération.

Je remercie vivement les évaluateurs de ce mémoire, Professeur Ismail Biskri et Mhamed Mesfioui pour leur disponibilité. Leurs observations ont joué un rôle important dans l'amélioration de ce travail.

À ma tendre épouse Cécile Koso Onyumbe Mizonza pour ton assistance et aussi pour ta patience et ta tolérance dont tu as fait preuve.

À mes enfants Lezain Eudenh Mizonza, Merveilleuse Mizonza, Glorieuse Mizonza et Gracieuse Mizonza sans oublier mes parents Samson Mizonza et Thérèse Mboya pour votre réconfort et votre amour à mon égard.

Que Ouedraogo Abdoulaye trouve ici mes remerciements les plus distingués pour ta franche collaboration de même qu'à mon frère Kevin Landry Tafeu pour avoir passé ensemble avec moi des moments aussi difficiles.

Mes remerciements les plus distingués s'adressent à mes chers collègues et amis du laboratoire LAMIA pour la collaboration dont je cite Toffa Zidane, Dipa Diallo, Ines Yahia, …

Je tiens à témoigner toute ma reconnaissance au professeur Jean Marcel Mbikayi.

Je remercie l'ensemble de la communauté congolaise au Canada dont je cite particulièrement Jonathan, Francis, Stanislas, Julien, Moise, Marianne, Oscar, Patrick, Chaïda, Julie, Serge, Stanley, Simon, Mamie, Beatrice, …

À la mémoire de Martin Ekanda qui nous a laissés tôt, à sa femme Ruffine Luyeye Ekanda et au couple Emmanuel Matondo, trouver dans ce mémoire l'expression de ma profonde considération.

Dans l'impossibilité de citer tous les noms, nos sincères remerciements vont droit à tous ceux et celles, qui de près ou de loin, ont permis par leurs conseils, leurs compétences et leurs apports la réalisation de ce projet.

Dieudonné MIZONZA BANTIKO

RÉSUMÉ

Aujourd'hui, avec la multitude des types d'attaques que les objets connectés sont victimes, nous avons besoin des systèmes de détection d'intrusion très efficaces et puissants, capables d'augmenter la sécurité des réseaux. C'est dans ce contexte que nous mettons en place un système intelligent de détection d'intrusion qui combine les deux algorithmes : un algorithme pour la classification (règles d'association maximale) et l'autre pour la modélisation (régression logistique) afin de produire un modèle pour la sécurité des réseaux véhiculaires V2G avec de meilleurs résultats, fiables et plusieurs fois mieux que toutes les solutions à algorithme unique. Chacun des algorithmes joue un rôle précis dans le système avec comme finalité, la reconnaissance des vraies et des fausses attaques. Nous avons simulé une base de données (que nous appelons Dataset), qui a été produite avec l'outil MiniV2G et toute sa suite (Mininet, RiseV2G, Wireshark, CICFlowmeter). Elle a été utilisée pour tester notre modèle par un traitement qui commence par une analyse de corrélation de Pearson avec le logiciel SPSS avant d'appliquer le processus d'extraction des règles d'association maximales que nous avons effectué avec le logiciel YamBob.

L'utilisation des règles d'association maximale nous a permis d'élaguer les règles qui ne sont pas pertinentes, en conservant celles qui sont intéressantes et de visualiser la perte des informations pour aboutir à une classification. Trois mesures d'évaluation ont été utilisées avec des valeurs prises de façon aléatoire : le support, la confiance et le lift.

Nous avons développé ensuite un modèle utilisant la méthode statistique appelée régression logistique exécutée avec le logiciel R. Avec 70% des données utilisées pour l'entraînement et 30% des données pour le test, avec un seuil de classification de 0,5; nous avons obtenu un taux de prédiction de 0,69 qui dépasse largement le seuil avec un taux de prédiction en pourcentage de 69% qui prouve l'efficacité du modèle.

Le test de notre modèle qui s'est fait en trois (3) séries d'essais avant de valider les valeurs des métriques devrait nous produire un bon résultat. Ce modèle proposé et expérimenté peut-être employé avec d'autres bases de données, à la seule condition de respecter la procédure et les formats des données à utiliser pour éviter d'obtenir des résultats biaisés.

MOTS CLES

IDS (Système de détection d'intrusions), règles d'association, règles d'association maximales, dataset, régression logistique, V2G, corrélation, MiniV2G**.**

ABSTRACT

With the multitude of types of attacks that connected objects are victims of, we need very effective and powerful intrusion detection systems capable of increasing network security. It is in this context that we are implementing an intelligent intrusion detection system that combines the two algorithms: one algorithm for classification (maximum association rules) and the other for modeling (logistic regression) in order to produce a model for the security of V2G vehicular networks with better results, more reliable and several times better than all single algorithm solutions.

Each of the algorithms plays a specific role in the system with the purpose of recognizing real and fake attacks. We simulated a database (which we call Dataset), which was produced with the MiniV2G tool and its entire suite (Mininet, RiseV2G, Wireshark, CICFlowmeter).

It was used to test our model by a treatment that begins with a Pearson correlation analysis with SPSS software before applying the process of extracting the maximum association rules that we carried out with YamBob software.

The use of maximum association rules allowed us to prune the rules that are not relevant, keeping those that are interesting, to visualize the loss of information to arrive at a classification. Three (3) evaluation measures were used with values taken at random: support, confidence and lift.

So, we proposed a model with the statistical method called logistic regression executed with the R software. With 70% of the data used for training and 30% of the data for the test, with a classification threshold of 0.5; we obtained a prediction rate of 0.69% which greatly exceeds the threshold with a percentage prediction rate of 69% which proves the effectiveness of the model.

The test of our model was done in three (3) series of tests before validating the values of the metrics should produce a good result. This model can be used with other databases, on the sole condition of respecting the procedure and the data formats to be used to avoid obtaining biased results.

KEYWORDS

IDS (Intrusion detection system), association rules, maximal association rules, dataset, logistic regression, V2G, correlation, MiniV2G.

TABLE DES MATIERES

LISTE DES TABLEAUX

LISTE DES FIGURES

SIGLES

IDS: Intrusion Detection System

V2G: Vehicle-To-Grid

VE: Electric Vehicle

IPS: Intrusion Prevention System

OSI: Open Systems Interconnection

ISO: International Organisation for Standardisation

CO2: Carbon Dioxide

EP: Electricity Provider

EVCC: Electric Vehicle Communication Controller

EVSE: Electric Vehicle Supply Equipment

ECU: Electronic Control Unit

SECC : Supply Equipment Communication Controller

DSO: Distribution System Operator

BC : Base de Connaissance

IDWG: Intrusion Detection exchange format Working Group

TLS : Transport Layer Security

OEM: Original Equipment Manufacturer

HIDS: Host-based Intrusion Detection System

NIDS : Network-based Intrusion Detection

RSU: Road Side Unit

DOS: Denial Of service

ABIDS : Application-Based Intrusion Detection System

EIM : External Identification Means

BMS: Battery Management System

DCH: Demand Clearing House

CA : Certificate Authority

PHEV: Plug-in Hybrid Electric Vehicle

TCP- IP : Transmission Control Protocol -Internet Protocol

V2GTP : Vehicle-To-Grid Transfer Protocol

UDP : User Datagram Protocol

XML Extensible Markup Language

PLC : Power Line Communication

PEV : Plug-in Electric Vehicles

BEV : Battery Electric Vehicles

PHEV : Plug-in Hybrids

CEI : Commission électrotechnique internationale

RFC : Remote Function Call

IETF : Internet Engineering Task Force

ICMP : Interchange Message Message Protocol

TLS : Transport Layer Security

SDP : Session Description Protocol

DDOS : Distributed Denial of Service

DOS : Denial of Service

SYN : synchronization

ACK : acknowledge

HTTP : Hypertext Transfer Protocol

PHP : Hypertext Preprocessor

SMS : Short Message System

MitM : Man in the Middle

TELNET : Telecommunication Network

SNMP : Simple Network Management Protocol

WEP : Wired Equivalent Privacy

WPA : Wireless Protected Access

HIDS: Host Based Intrusion Detection System

NIDS : Network Based Intrusion Detection System

NIPS : Network Based Prevention Detection System

HIPS : Host Based Intrusion Prevention System

WIPS : Wereless Intrusion Prevention System

PCAP : Packet Capture

LR : Logistic regression

CIC : Canadian Institute for CyberSecurity

V2H : Vehicle-To-Home :

V2B : Vehicle-To-Building :

ECDH : Elliptic Curve DiffieHellman

TTL : Time-to-live

SPSS : Statistical package for the social sciences

CHAPITRE 1 : INTRODUCTION GÉNÉRALE

La détection et la prévention des anomalies figurent parmi le principal objectif tant pour de nombreux chercheurs que pour des entreprises en raison de son potentiel à détecter de nouvelles attaques dans un système. Cependant, son adoption dans les applications du monde réel a été entravée en raison de la complexité de son système qui nécessite une quantité substantielle de tests, d'évaluation et de réglage avant son déploiement. L'exécution des systèmes de détection d'intrusion sur des véritables traces de réseau étiquetées avec un ensemble complet d'intrusions et de comportements anormaux est la méthodologie la plus idéaliste pour les tests et l'évaluation dans l'espoir de trouver une solution à ce problème d'intrusions. **[1]**

Les systèmes de détection et de prévention d'intrusions sont devenus très indispensables dans toutes les plateformes et surtout lors de l'implémentation des solutions de sécurité opérationnelle.

1.1. Contexte et objectifs

Les problèmes liés à la sécurité sont souvent complexes dans la mesure où la sécurité n'a jamais atteint 100%. Plusieurs solutions existent pour mettre à l'abri les constituants des systèmes informatiques en générale, et des réseaux publics en particulier qui sont toujours exposés aux risques d'attaques de pirates.

Le fait de ne pas atteindre un niveau de sécurité maximal est causé par l'utilisation de composants de la sécurité du réseau qui présentent toujours une vulnérabilité, voilà pourquoi plusieurs solutions sont proposées et déployées pour tenter de renforcer les composants de sécurité existants dans le souci de mettre fin à ses faiblesses. Parmi ces solutions nous pouvons citer celles qui consistent à regrouper des IDS (Systèmes de détection d'intrusion) pour créer un autre mur de protection permettant de protéger les systèmes et d'identifier les intrusions au-delà du pare-feu, par exemple.

Malgré l'ajout d'un mûr supplémentaire dans les systèmes informatiques (surtout ceux qui utilisent les réseaux publics), le niveau de vulnérabilité est toujours présent. Au lieu de continuer à ajouter du matériel pour surmonter les insuffisances des différents systèmes de sécurité existants, il est plutôt nécessaire d'optimiser les systèmes existants avec des approches systématiques afin d'arriver à générer des ensembles de données capables d'analyser, de tester et d'évaluer les systèmes de détection d'intrusion, en mettant l'accent sur les méthodes de détection d'anomalies en réseau [**1].**

Parmi les solutions d'optimisation des systèmes de détection d'intrusion existantes, nous avons choisi dans le cadre de ce mémoire une approche basée sur les règles d'association maximales et la régression logistique afin d'améliorer significativement les systèmes de détection d'intrusion dans les réseaux véhiculaires V2G (Vehicle-To-Grid) et plus précisément entre le véhicule électrique et la borne de recharge.

Étant donné que tous les systèmes de détection d'intrusion qui donnent des faux positifs et de faibles taux de contre-mesures d'anomalies peuvent avoir un impact très négatif dans le réseau, nous avons pensé dans le cadre de ce travail à résoudre les problèmes de limites que présentent certains systèmes de détection d'intrusion en se basant sur la détection d'anomalies, sur l'exploration des règles d'association maximales et de la régression logistique afin de faire le choix des vraies attaques à détecter et les fausses attaques à archiver.

Ainsi, avec notre solution, le risque sera faible car le niveau de vulnérabilité sera réduit de même que le nombre de menaces puisque nous allons écarter toutes les fausses attaques, et de ce fait, le risque de tomber sur une contre mesure inappropriée sera largement réduit.

Notre solution pour les réseaux V2G poursuit plusieurs objectifs spécifiques comme :

1. La détection de tous les trafics à l'entrée de nos réseaux véhiculaires V2G au niveau des bornes,
2. Le blocage automatique des activités anormales ou suspectes donc non autorisées,
3. La surveillance des toutes les activités des réseaux V2G, l'analyse de ses configurations contre les vulnérabilités,
4. L'analyse de l'intégrité des données, la confidentialité …
5. L'émission des alertes en cas de détection de vraies attaques
6. L'archivage des fausses attaques.

1.2. Problématique et hypothèse

Comme nous le savons, la sécurité dans les réseaux informatiques en général et dans les réseaux V2G en particulier est nécessaire donc, un système de détection d'intrusions est plus que crucial. Il permet à ce que toutes les activités anormales ou suspectes soient détectées et au préalable réparé. Dès que ces activités sont détectées, une alerte est souvent déclenchée pour signaler une anomalie dans le système.

Les systèmes de détection d'intrusions génèrent une énorme quantité d'alertes où la plupart d'entre elles sont réelles alors que d'autres ne le sont pas (fausses alertes) et parfois ce sont des alertes redondantes. Les fausses alertes créent un grave problème surtout pour les systèmes de détection d'intrusion **[2].** Ce comportement des systèmes de détection d'intrusion est dû aux limites ou vulnérabilités que présentent les équipements utilisés pour jouer ce rôle.

Comme tous les systèmes de sécurité ne sont pas capables de protéger un système à 100% et contiennent toujours un niveau de vulnérabilité, nous pensons que pour répondre à cette problématique sur les insuffisances que présente la sécurité, sa politique doit être définie en fonction du système que l'on souhaite sécuriser et des objectifs que l'on souhaite atteindre. Cette politique exprime les propriétés de confidentialité, d'intégrité, de la non-répudiation et de la disponibilité qu'ils doivent être respectées afin de garantir sa sécurité. **[3]**

Pour faire respecter ces propriétés, des mécanismes préventifs, notamment de contrôle d'accès, des IDS sont mises en œuvre sur les systèmes d'information. Ces mécanismes ont accès à tout ou à une partie de la politique de sécurité et devraient être capables d'empêcher de manière préventive toute action qui aboutirait à la violation d'une des propriétés qu'elle exprime. **[3]**

C'est dans ce contexte que nous avons pensé proposer dans ce mémoire une solution qui sera capable de minimiser les fausses alertes, augmenter le taux de détection et de filtrer les attaques afin de bloquer les vraies et supprimer ou archiver les fausses. Nous pensons qu'avec les règles d'association maximales couplées à la régression logistique, notre solution va résoudre ces problèmes de sécurité auxquels le réseau véhiculaire V2G est confronté.

1.3. Méthodologie et outils

Le déploiement de la solution impliquera plusieurs méthodes et outils. Les règles d'association maximales seront utilisées pour connaitre les relations des messages ayant une importance dans la base de données qui est considérée comme attaque dans le réseau V2G. Elles permettront de dégager des relations intéressantes dans le corpus en enlevant les relations les moins intéressantes qui ne peuvent pas être capturées par des règles d'associations ordinaires. Pour ce faire, le logiciel YamBob développé par ABDOULAYE OUEDRAOGO **[4]** sera utilisé pour extraire les règles d'association les plus importantes et élagué celles qui ne nous serviront pas.

Pour mieux étudier la qualité des règles et éliminer les règles d'association redondantes, que nous allons ajouter aux règles d'association, une solution statistique comme la régression logistique

sera utilisée. Le modèle de la régression logistique sera utilisé pour chercher à établir une relation entre la variable expliquée et les variables explicatives des données venant du traitement des règles d'association maximales de notre dataset afin de ressortir les variables significatives en relation avec la variable expliquée qui est considérée comme attaque et ainsi elle ressortira le modèle final. L'outil à exploiter pour le traitement de ses données est le logiciel Rstudio [**4**]. Cette méthode et le logiciel nous permettront de terminer la modélisation de l'optimisation de notre système. Nous utiliserons le simulateur MiniV2G pour générer notre dataset qui constituera notre base de données de départ. L'analyse de la corrélation de Pearson sera la méthodologie utilisée pour réduire notre dataset.

1.4. Contribution

Notre travail de mémoire vise à améliorer les systèmes de détection et de prévention d'intrusions (IDS et IPS) dans le réseau V2G. Notre souci qui consiste à chercher cette amélioration se justifie par les raisons suivantes:

a) L'augmentation du nombre des attaques dont les différents réseaux (surtout les réseaux publics) sont victimes. Ces attaques nous obligent à trouver des solutions capables d'améliorer les systèmes de défense ou de sécurité;
b) Les faiblesses que présentent les actuels systèmes de détection d'intrusion qui génèrent beaucoup de fausses alertes et aussi des alertes redondantes;
c) Difficulté des administrateurs des systèmes à gérer leur réseau avec une multitude d'alertes et une haute vigilance dû à la vulnérabilité que présente le système de sécurité;
d) Le manque de systèmes de sécurité, moins encore des IDS qui sont capables de résoudre les problèmes liés à la reconnaissance des vraies attaques et alertes dans le réseau V2G;
e) L'utilisation des méthodologies déjà existantes et connues par les attaquants dans l'optimisation des IDS. Parfois, les données des dataset qui sont utilisées dans les tests des différentes solutions ne représentent pas tous les cas d'attaques dont ils sont victimes.

Tenant compte des raisons évoquées ci-dessus, notre contribution à travers ce travail de recherche est de proposer une solution fiable qui permet d'optimiser les IDS avec des méthodes et des outils adaptés.

Notre contribution se résume à :

a) L'utilisation des méthodologies statistiques, efficaces comme la statistique descriptive et la régression logistique pour le traitement du dataset pour obtenir une solution modélisée, fiable, performante et très réactive face aux attaques;

b) L'utilisation des règles d'association maximales afin de trier les vraies et les fausses attaques puis les alertes;
c) L'utilisation des outils informatiques afin d'automatiser les activités de notre solution et obtenir ainsi des résultats fiables;
d) L'implémentation d'une solution des IDS qui répond aux objectifs de la sécurité et aux besoins du V2G selon les normes de l'ISO (International Organisation for Standardisation), qui peut aussi être adaptée dans d'autres réseaux tant publics que privés.

1.5. Structure du mémoire

Le mémoire est structuré en plusieurs chapitres. Il commence par le chapitre Introduction générale qui donne un aperçu général du travail. Le deuxième chapitre nous permettra de faire une présentation du réseau V2G et de la norme ISO15118 afin de comprendre le réseau dans lequel nous allons déployer notre solution. Quant au troisième chapitre, il aborde les notions de bases des systèmes de détection d'intrusions pendant que le quatrième chapitre nous permettra de faire une étude précise des règles d'association, des règles d'association maximale et de la régression logistique. Au cinquième chapitre, nous allons présenter une revue de la littérature sur les recherches récentes et voir les résultats obtenus dans les systèmes de détection d'intrusion. Dans le sixième chapitre, nous allons produire notre dataset puis proposer notre modèle d'optimisation des systèmes de détection d'intrusion dans les réseaux V2G. Au chapitre 7, nous présenterons une expérimentation des résultats. Enfin, nous présenterons la conclusion générale qui vient confirmer les résultats obtenus et projeter les perspectives d'avenir.

CHAPITRE 2 : RÉSEAU VÉHICULAIRE V2G ET LA NORME ISO15118

2.1. Introduction

Il est important de reconnaitre que l'humanité à intérêt de bien protéger sa planète pour espérer vivre dans un bon environnement à pollution presque nulle afin d'éviter ou au moins réduire le réchauffement climatique. Seulement, depuis un certain temps, de développement technologique oublie les intérêts humanitaires et contribue à un pourcentage important à la pollution de la planète terre.

L'industrie de l'automobile n'est pas épargnée avec la fabrication des véhicules à combustible (essence, gasoil) qui dégagent beaucoup de gaz carbonique (CO2) et qui polluent l'environnement.

C'est dans ce contexte que le domaine de l'automobile a pensé faire la migration des véhicules à combustible vers les véhicules électriques pour protéger notre planète. Cette migration implique des préalables comme la mise en œuvre des normes capables de règlementer cette technologie.

Dans ce chapitre, nous allons non seulement présenter le réseau véhiculaire V2G (Vehicle-To-Grid), mais également les deux premières parties de la norme ISO 15118 dont :

1. International standard ISO 15118-1(première partie) ;
2. International standard ISO 15118-2 (deuxième partie).

2.2. Réseau V2G

2.2.1. Présentation

Le réseau Vehicle-To-Grid (V2G) décrit un système dans lequel l'ensemble constitué des véhicules électriques rechargeables (PEV : Plug-in Electric Vehicles), des voitures électriques et des hybrides rechargeables, sont capables de communiquer avec le réseau électrique afin de faciliter les services de réponse à la demande en chargeant ou en déchargeant de l'énergie. Il se présente deux cas de figure :

1. Au cas où les PEV (Plug-in Electric Vehicles) font la recharge à partir d'un réseau électrique, alors l'énergie est emmagasinée dans leurs batteries de stockage;
2. Au cas où les PEV (Plug-in Electric Vehicles) font la décharge vers un réseau électrique, c'est maintenant l'énergie qui est stockée dans leurs batteries qui sera transférée vers le

réseau électrique afin de stabiliser l'équilibre dans la demande d'énergie, car l'échange de l'énergie se fait en bidirectionnel. **[5]**

2.2.2. Concepts

Nous allons dans ce qui suit définir quelques concepts qui font partir du réseau V2G tel que la norme ISO 15118-1 les a définis:

1. **Système de gestion de batterie (BMS)** : c'est un dispositif électronique qui contrôle ou gère les fonctions électriques et thermiques du système de batterie et qui assure la communication entre le système de batterie et les autres contrôleurs du véhicule ;
2. **Certificat** : c'est un document électronique qui utilise une signature numérique pour lier une clé publique à une identité ;
3. **Centre d'échange à la demande (DCH)** : c'est une entité de négociation de la grille fournissant des informations sur la charge de la grille ;
4. **Opérateur E-Mobility** : c'est une entité avec laquelle le client a un contrat pour tous les services liés à l'opération EV ;
5. **Fournisseur d'électricité EP** : c'est un organisme d'acteurs secondaires pour fournir de l'électricité ;
6. **Véhicule électrique EV** : c'est tout véhicule propulsé par un moteur électrique puisant du courant dans une batterie de stockage rechargeable ou provenant d'autres dispositifs de stockage d'énergie portables (rechargeables, utilisant l'énergie d'une source externe au véhicule comme un service électrique résidentiel ou public), fabriqué principalement pour être utilisé sur des réseaux publics : rues, routes ou autoroutes ;
7. **Contrôleur de communication de véhicule électrique (EVCC)** : c'est un système embarqué, à l'intérieur du véhicule, qui met en œuvre la communication entre le véhicule et le SECC afin de supporter des fonctions spécifiques ;
8. **Équipement de fourniture de véhicules électriques (EVSE)** : c'est un ensemble constitué des conducteurs, les phases, les conducteurs de neutre et de protection, les coupleurs EV, attaches prises, ainsi que tous les autres accessoires, appareils, prises de courant ou appareils installés spécialement dans le but de fournir de l'énergie du câblage des locaux au VE et de permettre la communication entre eux si nécessaire ;
9. **Unité de contrôle électronique (ÉCU)** : c'est une unité fournissant des informations sur le véhicule ;
10. **Opérateur EVSE** : c'est un acteur chargé de la gestion et de la maintenance du point de charge ;

11. **Contrôleur de communication d'équipement d'approvisionnement (SECC)** : c'est une entité qui met en œuvre la communication avec un ou plusieurs EVCC conformément à ISO 15118-2 et qui peut être en mesure d'interagir avec des acteurs secondaires. **[6]**

2.2.3. Architecture

La Figure 2.1 ci-dessous, décrit l'architecture générale d'un système V2G en définissant les interactions, la production et la transmission de l'énergie, les consommateurs d'énergie et les utilisateurs des PEV. **[5]**

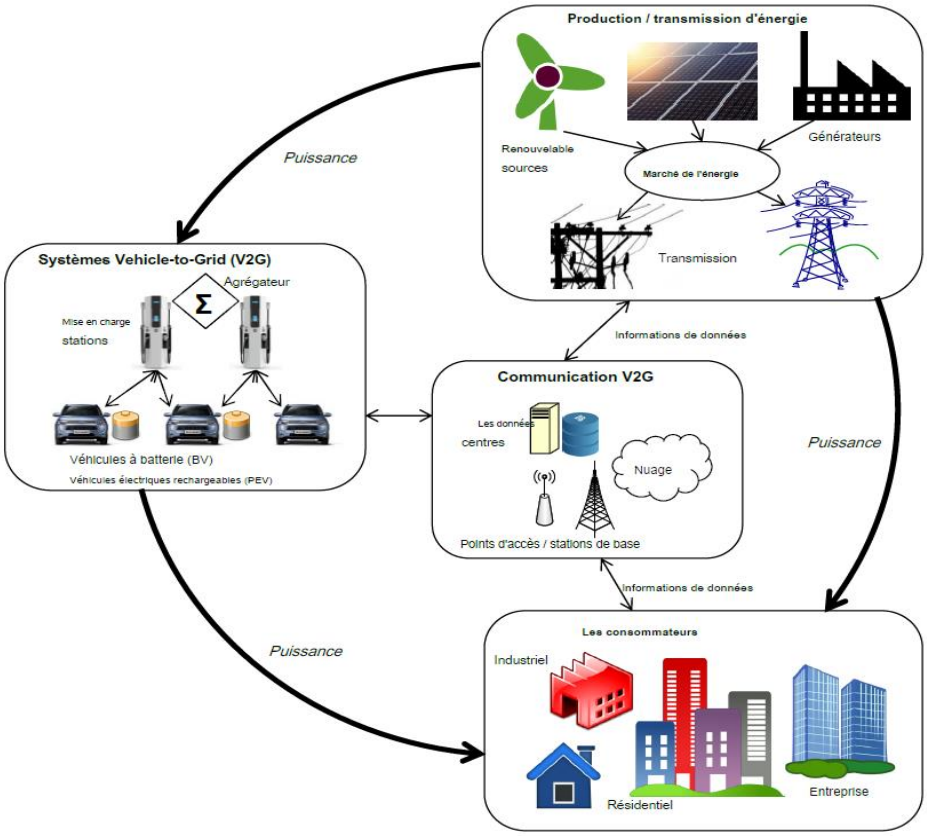

Figure 2.1. : Architecture du réseau V2G [5]

Comme on peut le voir dans la **Figure 2.1,** il existe quatre parties ayant chacune un rôle bien précis dans l'architecture du réseau V2G : 1- la production et la transmission de l'énergie, 2- le système V2G, 3- la communication V2G et 4- les consommateurs.

2.2.3.1. Production et la transmission de l'énergie

Cette partie constitue les systèmes électriques dans lesquels se trouvent des générateurs conventionnels d'énergie, des sources d'énergie renouvelable et des dispositifs pour la transmission de

l'énergie. Elle est chargée de fournir de l'énergie électrique au réseau V2G et aux consommateurs comme les immeubles, les entreprises, les commerciaux … Voilà pourquoi, elle a plusieurs sources d'énergie, car cette partie ne peut jamais manquer d'énergie à fournir aux clients. **[5]**

2.2.3.2. Système V2G

Ce système est constitué de plusieurs bornes de recharges et des PEV qui se connectent au réseau électrique via ces bornes ou sur les agrégateurs publics et privés afin de se recharger ou de charger les bornes, car le V2G est au même moment un consommateur et organe de stockage d'énergie. C'est grâce aux agrégateurs qui contrôlent et optimisent les flux d'énergie qui s'effectue dans cette partie que la médiation est assurée entre le réseau électrique et les systèmes V2G. **[5]**

2.2.3.3. Communication V2G

La communication V2G permet l'échange de données entre les systèmes d'alimentation, les consommateurs d'énergie et les systèmes V2G. Pour jouer ce rôle, elle se compose d'une infrastructure de communication (réseaux sans fil) et d'installations de traitement (cloud computing et centre de données). Avec l'infrastructure de communication V2G, les gestionnaires du réseau électrique peuvent collecter les données nécessaires auprès des systèmes et des consommateurs V2G, puis optimiser efficacement la production d'électricité et les services auxiliaires des PEV.**[5]**

2.2.3.4. Consommateurs

Les utilisateurs de PEV peuvent conclure un accord / contrat à long terme avec l'opérateur V2G pour rendre la charge et la décharge plus prévisibles. Par exemple, l'opérateur peut proposer un service de maintenance de la batterie en échange des utilisateurs PEV acceptants de charger et décharger la batterie pour répondre aux exigences du V2G **[5]**

2.2.4. Recharge des VE

Le système de la recharge d'un véhicule électrique est lié à plusieurs exigences qui doivent être pris en compte. Il y a ceux liés à la prise en compte de l'énergie électrique disponible dans le système, à l'optimisation du coût de la recharge, de la disponibilité du véhicule à recharger, et le maintien des performances des batteries lors de la recharge. **[7]**

Pour cela, il nous faut des bornes connectées à la grille et utilisant intelligemment les ressources. Cette intelligence est rendue possible par l'échange qui se fait entre le contrôleur du

chargeur de la borne et le calculateur principal du véhicule sans oublier la coordination des différents organes qui sont capables d'influencer le processus de la recharge ou de la décharge. On peut donc citer : Battery Management System (BMS), le réseau électrique, l'utilisateur, la météo comme on peut le voir dans la **Figure 2.2** ci-dessous. **[7]**

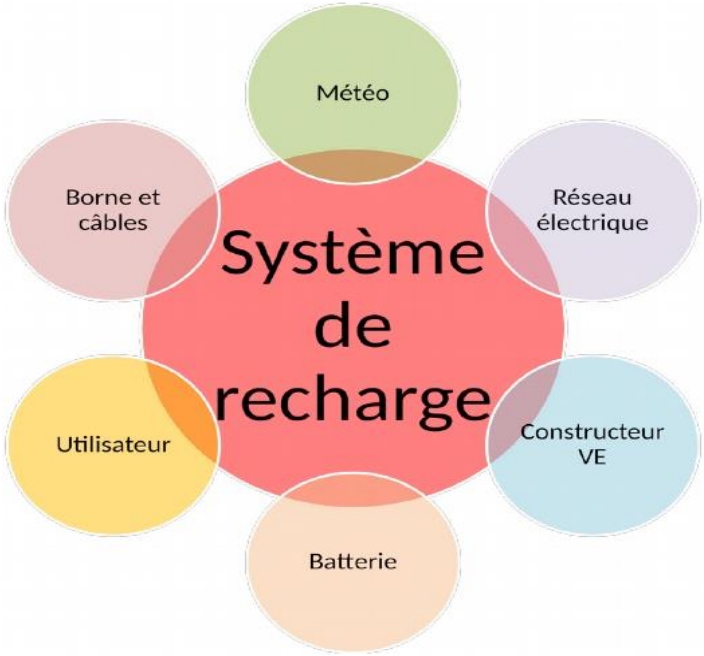

Figure 2.2. : Contraintes à respecter dans le cas d'une recharge planifiée [7]

2.3. Présentation de la norme ISO 15118

La norme ISO 15118 est une norme qui enrichit considérablement la communication entre un véhicule électrique et une infrastructure de recharge. Elle complète l'architecture informatique de l'écosystème de la recharge des véhicules électriques, permettant à celle-ci d'être plus simple, plus sécurisée et d'apporter de nouveaux services aux conducteurs :

1. Le Plug and Charge qui permet d'identifier automatiquement le contrat de services de l'utilisateur par simple branchement du câble de recharge entre le véhicule et la borne, avec un haut niveau de sécurité informatique et une expérience utilisateur simplifiée ;
2. La gestion intelligente de la recharge (smart charging) qui permet de programmer un planning de recharge, négocié entre la borne et le véhicule et optimisé selon leurs contraintes techniques, les besoins et exigences du conducteur et les contraintes électriques du réseau ;
3. La bidirectionnalité de la recharge qui permet de négocier et d'optimiser, via la borne, la réinjection d'électricité stockée dans le véhicule à la maison (vehicle-to-home : V2H), au bâtiment (vehicle-to-building : V2B) ou au réseau (vehicle-to-grid : V2G) ;
4. D'autres services pourront être conçus et apportés aux conducteurs, par exemple pour la recharge sans manipulation de câble (recharge inductive ou automatisée). **[8]**

C’est une norme internationale qui spécifie le protocole de communication numérique qu'un véhicule électrique (VE) et une station de charge doivent utiliser pour recharger la batterie haute tension de ce dernier. Elle couvre tous les cas d'utilisation liés à la charge à travers le monde. **[8]**

C’est grâce à cette norme qui a été élaborée en coopération avec le CEI TC 69 (Véhicules routiers électriques et chariots de manutention électriques) que nous avons la possibilité de charger les Véhicules Electriques de manière plus conviviale, sécurisée et plus pratique sur le réseau V2G.

L'ISO 15118 comprend plusieurs parties qui sont présentées sous forme de titre général Véhicules routiers - Véhicule au réseau interface de communication. Ces parties sont :

1. Partie 1 : Informations générales et définition de cas d'utilisation ;
2. Partie 2 : Exigences du réseau et du protocole d'application ;
3. Partie 3 : Spécifications de la couche physique et de la couche liaison de données
   Deux autres parties sont en cours de préparation ;
4. Partie 4 : Test de conformité de réseau et de protocole d'application ;
5. Partie 5 : Essai de conformité de la couche physique et de la couche liaison de données **[6]**.

À ces 5 parties s’ajoutent encore 3 autres pour donner actuellement 8 parties de cette norme.

2.3.1. Présentation de la norme ISO 15118-1

Cette première partie de l’ISO 15118 était sortie en deux éditions dont une première le 15 avril 2013 et une deuxième qui est une version corrigée parue le 1er octobre 2013. Cette norme a dans l’ensemble expliqué les objectifs généraux en définissant les termes et ses cas d'utilisation du réseau véhiculaire V2G. **[6]**

2.3.1.1. Portée

Elle présente de façon générale la norme ISO 15118-1 qui spécifie la communication entre les véhicules électriques (VE), leurs batteries électriques et véhicules électriques hybrides rechargeables et les équipements de fourniture de véhicules électriques (EVSE), l’interaction de différents composants qui composent ce système. Elle spécifie et définit aussi les termes utilisés dans le V2G, les exigences, les cas d’utilisation, la facturation, le paiement, les catégories des véhicules concernés (M et N).

2.3.1.2. Références normatives

Elles présentent les normes qui ont contribué à la concrétisation de cette norme en citant nommément le CEI 60050 (Commission Électrotechnique Internationale) qui spécifie le vocabulaire électrotechnique international, le CEI 61851-1 pour le Système de charge conductrice pour véhicule électrique partie 1, 2 et 3.

2.3.1.3. Exigences

Elles s'articulent autour des aspects suivants :

- **Concept de communication** : les échanges d'informations ou la communication à haut niveau n'ont lieu que si le VE et le EVESE sont équipés des appareils de communication de haut niveau.

La communication de haut niveau est utilisée seulement pour activer les fonctionnalités de l'identification, le paiement, le nivellement de la charge et les services à valeur ajoutée.

- **Considérations générales** : elles abordent plus les aspects liés à la sécurité au niveau de la communication en explicitant les cas pour lesquels interviennent la confidentialité, la cryptographie, la protection contre toute modification ou imitation (piratage), la mesurabilité et la facturation.
- **Exigences spécifiques à l'utilisateur** : elles donnent les garanties à l'utilisateur sur la qualité et la fiabilité de ses données. Elles spécifient la fiabilité de ses données, la disponibilité du système, le traitement et la signalisation des erreurs, la protection et la confidentialité de ses données personnelles.
- **Exigences spécifiques aux constructeurs** : Un programme de facturation est calculé soit par un acteur secondaire, l'EVSE ou par l'EV, sur la base des informations fournies par l'utilisateur, le point de charge et le réseau d'énergie, puis il est transféré vers le réseau pour permettre la planification d'autres véhicules électriques. Mais, elles spécifient aussi les possibilités pour lesquelles les informations du client peuvent être modifiées. Par exemple : à la production du VE, à la livraison du VE, lorsque le contrat énergétique est modifié par le client …
- **Exigences spécifiques aux utilitaires** : ici, la norme spécifie les conditions de la limitation de puissance pour le contrôle du réseau ou de l'énergie locale, de la limitation du courant pour la protection des EVSE, l'autorisation des services de facturation et de la rénovation **[6].**

2.3.1.4. Acteurs

La norme ISO 15118-1 a aussi défini les acteurs qui interviennent dans la technologie V2G et leur fonction. Deux catégories d'acteurs sont reconnues : acteur primaire et acteur secondaire.

Les acteurs primaires sont directement impliqués dans le processus de chargement (VE, utilisateur, EVSE, SECC, …) et les acteurs secondaires qui sont le DSO (Distribution System Operator), le Provider (Fournisseur), opérateur E-mobile … **[6]**

2.3.1.5. Éléments de cas d'utilisation

Les cas d'utilisation qui sont abordés dans cette partie de la norme concernent plus le processus de charge du VE. Ce processus part du début de la recharge jusqu'à la fin de la recharge en spécifiant les différents scénarios qui sont divisés en 8 groupes fonctionnels ayant chacun plusieurs cas d'utilisation élémentaires pour permettre la classification des éléments élémentaires :

1. Début du processus de charge : lancement du processus entre véhicules et EVSE après le contrôle physique plug-in du véhicule.
2. Configuration de la communication : établie l'association et la connexion pertinente entre EVCC et SECC ;
3. Gestion des certificats : tout ce qui concerne l'utilisation des certificats ;
4. Identification et autorisation : Les méthodes d'identification et d'autorisation ;
5. Fixation de la cible et ordonnancement de la facturation : Les informations requises de la part du VE ainsi que de la SECC et l'acteur secondaire pour démarrer le processus de charge et la charge ;
6. Contrôle et reprogrammation de la facturation : Les éléments au cours du processus de facturation ;
7. Services à valeur ajoutée : Les éléments non nécessaires à la recharge pure des véhicules électriques ;
8. Processus de fin de charge : Les déclencheurs signalant la fin du processus de charge.

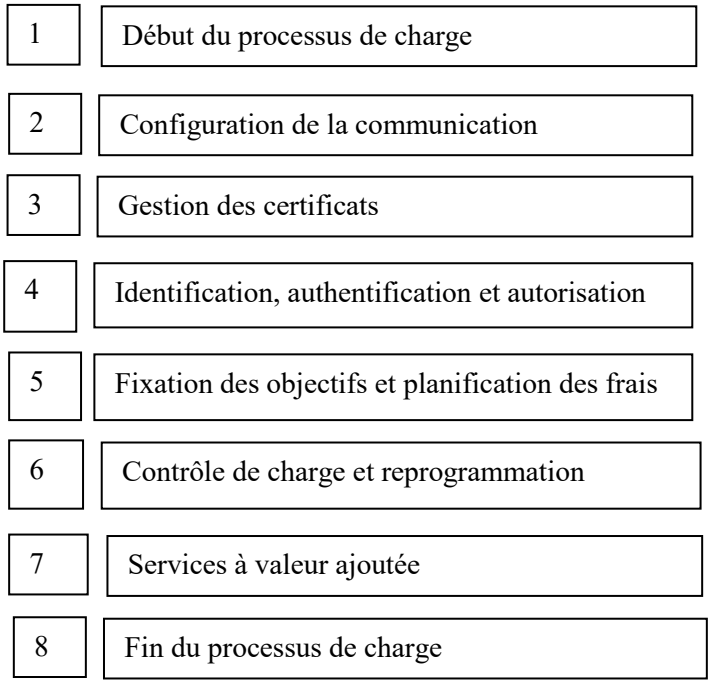

Figure 2.3. : Groupes de fonctions des éléments de cas d'utilisation [6]

Le scénario spécifié dessus définit le principe de la facturation, l'autorisation de la recharge, la modalité de paiement …

2.3.1.6. Forces de la norme

Bien que plusieurs cas d'utilisation ne soient pas décrits dans cette partie de la norme ISO 15118-1, mais qui sont décrits à la deuxième édition de l'ISO 15118-1, plusieurs points donnent une force à cette première partie de la norme :

1. Définition des termes et des scénarios de communication ;
2. Identification automatisée et autorisation de la recharge par la fonctionnalité plug and charge ;
3. Identification du VE ;
4. Chargement filaire ;
5. Gestion de charge (charge intelligente) pour tous les modes de charge ;
6. Possibilité au chargement sans fil ;
7. Possibilité de transfert de puissance bidirectionnel, qui permet au VE de fournir de l'énergie au réseau ;
8. Traitement des erreurs.

2.3.1.7. Analyse critique de la norme ISO 15118-1

Malgré les forces et les informations générales fournies par cette première partie de la norme, elle présente néanmoins quelques faiblesses telles que :

- La norme ISO 15118-1 n'a pas donné des informations sur le problème lié au transfert des données entre le VE, l'installation du local et du réseau ;
- Elle n'a pas précisé le niveau d'intervention et la contribution des autres normes qui ont été associées à elle, telle que les RFC de l'IETF ;
- À cela s'ajoute le problème des tests de conformité qui sont spécifiés dans cette norme et qui malheureusement n'incluent pas l'évaluation du rendement, de la robustesse ou encore de la fiabilité d'une implémentation ;
- Elle parle des acteurs primaires et secondaires, mais elle n'inclut pas des éléments permettant l'échange de données entre ces acteurs qui sont renvoyés à la deuxième partie.

2.3.2. Norme ISO 15118-2

Cette deuxième partie de la norme ISO 15118-2, est dénommée : Spécifications du réseau et du protocole d'application. Elle vient compléter certaines faiblesses de la première partie que nous venons de résumer ci-dessus. Elle est apparue en 2014 et détermine la communication entre un point de charge et un véhicule électrique (VE) et elle constitue la partie centrale de cette norme. **[8]**

2.3.2.1. Portée

La norme ISO 15118-2 spécifie la communication entre les véhicules électriques à batterie (BEV) et les véhicules hybrides rechargeables, les véhicules électriques (PHEV) et l'équipement d'approvisionnement en véhicules électriques. Cette partie prend aussi en compte le transfert de l'Énergie d'un EVSE à un VE, le format des données et elle définit le mode d'accès aux services de la couche 2 via un terminal intelligent.

Elle reprend aussi certains cas d'utilisation qui étaient déjà développés dans la première partie, spécifie la détection de VE qui se trouvent dans le réseau de communication, et active le protocole IP entre le EVCC et SECC et les messages de la couche application. **[9]**

2.3.2.2. Références normatives

Plusieurs références (datées et non datées) ont été utilisées dans cette partie de la norme dont on peut citer : ISO 3166-1, ISO 15118-1, CEI 61851-1, CEI 61851-22, IEC CDV 61851-23, CEI 62196, IETF RFC 768, IETF RFC 793, IETF RFC 1981, …

2.3.2.3. Conventions

Elles constituent un ensemble des règles adoptées et écrites dans le but de formaliser la réalité de la technologie du réseau véhiculaire V2G. Ces conventions concernent :

1. **Définition des services OSI** (Open Systems Interconnection) : les conventions sont définies par rapport à l'architecture en couche du modèle OSI. Elle décrit ses exigences applicables aux couches allant de 3 à 7 du modèle OSI ;
2. **Structure des exigences** : les exigences ou conventions de cette partie sont décrites suivant une structure à numéro unique qui permet d'identifier chaque exigence incluse dans le document ;
3. **Utilisation des références RFC** (Remote Function Call) : ce sont des références liées à la RFC ;
4. **Notation utilisée pour les diagrammes de schéma XML (**Extensible Markup Language) : elle représente le format utilisé pour faire la description des messages V2G. **[9]**

2.3.2.4. Configuration de base pour la communication V2G

La présente partie de l'ISO 15118 décrit la réalisation des éléments de cas d'utilisation V2G définis dans la partie 1 de la norme, l'application du modèle OSI (OSI : Open Systems Interconnection) à l'ISO 15118 (ISO: International Organisation for Standardisation) et les interfaces entre les protocoles.

Dans cette norme sont décrites aussi les fondamentaux des services selon la syntaxe suivante : [initial de la couche] - [NOM]. [Type primitif] (liste de paramètres)
Où :

1. [initial de la couche] est l'un des sept suivants : [Physique, liaison de données, réseau, transport, session, présentation et application]
2. [NOM] est le nom de la primitive : EXEMPLE : CONNECT, DISCONNECT, DATA
3. [Type primitif] est l'un des quatre suivants : [demande, indication, réponse, confirmation]
4. (Liste de paramètres) inclut une liste de paramètres séparés par une virgule

2.3.2.5. Concept de sécurité

Dans cette partie de la norme est spécifié l'aspect sécurité de la technologie V2G puis un flux d'appel est utilisé pour expliquer la sécurité des informations. Il n'y a que les données de la sécurité qui doivent être mise en évidence, voilà pourquoi, on fournit un mécanisme de protection de base fondé sur le transport.

On peut utiliser le protocole TLS (Transport Layer Security) pour la communication et le transport entre EVCC et SECC ou non, car déjà les messages spécifiques sont protégés par le format du message XML. Si les données doivent être protégées sur le chemin ou vers un acteur secondaire, la protection doit durer plus longtemps. Le système est indépendant de tout mécanisme de protection supplémentaire aux niveaux inférieurs à la couche 3 du modèle en couche OSI, mais la prise en charge de TLS est obligatoire pour l'EVCC pour tous les modes d'identification à l'exception de l'identification en mode "EIM" avec Message Set EIM (External Identification Means) en charge CA (Certificate Authority) et Message Set EIM en charge DC.

Si le SECC ne prend pas en charge le protocole TLS, il peut y avoir un abandon des sessions de chargement avec certains véhicules électriques, dans la mesure où il est utilisé sous la responsabilité du VE d'accepter des sessions sans TLS. **[9]**

2.3.2.6. Certificat et gestion des clés

Il existe des certificats SECC utilisés dans la couche TLS pour permettre à EVCC d'authentifier le SECC et des certificats de contrat qui sont utilisés au niveau application pour s'authentifier auprès d'un SECC et/ou d'un acteur secondaire. Un certificat peut-être racine V2G ou éventuellement certificats de sous-autorité de certification qui certifient le SECC.

Il peut exister des certificats racines OEM (Original Equipment Manufacturer) et des certificats de fourniture OEM qui sont utilisés pour installer et mettre à jour les certificats de contrat.

Pour la fiabilité dans le transport des données, l'adressage, la gestion des messages erreurs et pour sa sécurité, le V2G s'appuie sur d'autres protocoles comme le TCP (Transmission Control Protocol), IP (Internet Protocol), ICMP (Interchange Message Message Protocol), UDP (User Datagram Protocol), TLS. Le SECC est obligatoirement configuré avec une adresse IP valide (de la version 6) pour permettre une connexion à un EVCC. Le mécanisme d'attribution de l'adresse IP pour le SECC n'a aucun impact sur la compatibilité et n'est pas défini par cette norme. Mais une fois

l'adresse IP attribuée, le SECC doit démarrer le serveur SDP (Session Description Protocol) du réseau V2G. **[10], [9]**

Le protocole de transfert V2G (V2GTP : Vehicle-To-Grid Transfer Protocol) est un protocole propre au V2G qui est utilisé pour la communication compacte de transfert des messages V2G entre deux entités V2GTP. Il consiste principalement en une définition d'en-tête et de charge utile permettant de préparer et de traiter efficacement les messages V2G. Le V2GTP est le protocole de transfert standard entre EVCC et SECC, mais peut également être utilisé pour la communication avec d'autres entités V2G prenant en charge le protocole V2GTP.

Toujours dans l'aspect sécurité, un mécanisme de cryptage avec des clés privées appartenant aux certificats de contrat doivent être protégé (c'est-à-dire chiffrées) lorsqu'elles sont distribuées à partir du serveur du SA à la EVCC. Chaque entité V2G doit disposer de mécanismes pour traiter l'échange de clés ECDH (Elliptic Curve DiffieHellman) et la clé privée correspondant au certificat de contrat ne doit être transmise que sous forme cryptée **[10], [9]**

En général, deux paires de mécanismes de sécurité sont prises en charge :

1. Authenticité et intégrité : génération de signature, vérification de signature ; la signature basée sur XML,
2. Confidentialité : cryptage, décryptage.

Les messages transmis disposent des en-têtes appelés « en-tête SDP : Session Description Protocol » dont le traitement est basé sur le traitement d'en-tête de message V2GTP.

La norme donne des spécificités sur les exigences dans toutes les couches du modèle OSI. **[9]**

2.3.2.7. Messages de la couche application

Les échanges des données qui sont effectués entre l'EVCC et SECC à la couche 7 sont basés sur l'architecture client-serveur où l'EVCC est considéré comme demandeur des services (client) dans la procédure de la facturation bien sûr et le SECC est le serveur qui répond aux requêtes formulées par le client EVCC.

Ces échanges comprennent deux ensembles de messages différents :

1. Messages de négociation de protocole de couche d'application V2G ;
2. Messages de la couche d'application V2G.

L'ordre de la transmission se fait en envoyant d'abord l'octet le plus significatif en premier lieu, puis l'octet le moins significatif en dernier lieu. La négociation peut se faire par Séquence de poignée de main ou par support. **[9]**

Un message V2G de la couche 7 est structuré en message V2G, entête de message V2G (qui contient des informations générales incluses dans tous les messages) et le corps du message V2G (contient des informations détaillées relatives à un message spécifique).

La partie de cette norme décrit aussi le mécanisme de traitement de la session de communication V2G, celui lié au paiement de la facture et de la gestion du certificat. **[9]**

2.3.2.8. Calendrier de communication V2G

L'échange de messages est basé sur deux catégories de minuterie : Message Timer : (surveille l'échange d'un message de demande et du message de réponse correspondant)

Et le temporisateur de séquence (surveille l'échange de plusieurs paires demande-réponse).

On retrouve dans la norme plusieurs types de synchronisations dont :

- Synchronisation des messages
- Synchronisation de séquence
- Synchronisation continue
- Synchronisation de la communication
- Synchronisation Cable Check
- Synchronisation de la précharge

Les Conditions requises pour la séquence des messages de demande-réponse ont été règlementées dans cette partie. Des identifiants uniques sont utilisés pour garantir que EVCC et SECC puissent se référer à une annexe SAS spécifique pendant la session de la communication V2G entière. **[9]**

2.3.2.9. Application de certificats

Dans le protocole de facturation, les types de certificats suivants sont utilisés :

- Certificats racines V2G : il s'agit de certificats racines globalement valides (de premier niveau). Ils sont utilisés pour vérifier l'authenticité des certificats. Les clés privées correspondantes sont en possession des autorités de certification racine respectives.
- Certificat racine d'opérateur de mobilité : la clé privée de ce type de certificat est utilisée pour signer les autres certificats
- Certificat de contrat : ce type de certificat est utilisé dans le cas d'utilisation Plug and Charge.
- Certificat SECC : ce type de certificat est utilisé pour authentifier le SECC auprès de l'EVCC.
- Certificat racine d'opération privé : un certificat racine d'opération privé est très similaire à un opérateur de mobilité.
- Certificat d'approvisionnement OEM : Ce type de certificat est individuel pour chaque véhicule.
- Certificat racine OEM : il est utilisé pour signer des certificats d'approvisionnement OEM. **[10], [9]**

L'installation du certificat se fait automatiquement via le protocole de facturation, il peut être envoyé de l'acteur secondaire au client sous forme de fichier et doit être installé dans le véhicule à l'aide d'une connexion en ligne ou de l'interface de diagnostic (dans un garage).

2.3.2.10. Forces de la norme

Bien que plusieurs cas d'utilisation ne soient décrits dans la norme ISO 15118-1 et de l'ISO 15118-2, plusieurs points donnent une force à cette deuxième partie de la norme dont on peut énumérer :

3. La définition des termes non définis dans la première partie,

4. La définition de certains mécanismes de sécurité basés sur les certificats, par exemple,

5. La collaboration avec d'autres protocoles,

6. La sécurité des messages …

2.3.2.11. Analyse critique de la norme ISO 15118-1

Malgré les informations fournies par cette deuxième partie de la norme, elle présente néanmoins quelques faiblesses telles que :

- La procédure de paiement et de la sécurité entre la borne de recharge et le VE qui n'ont pas été clairement définies,

- Elle parle de l'utilisation des adresses IPv6 sans bien spécifier son utilisation et la manière d'attribution des adresses IPv6,
- Le manque de certaines normes qui sont renvoyées à la partie 3.

2.4. Communication entre véhicule électrique et borne avec l'ISO/IEC 15118

La possibilité de communication entre le véhicule électrique et la borne de recharge en utilisant ISO/IEC 15118 permet de faire une bonne planification de la recharge en fonction d'une bonne négociation entre les besoins de l'utilisateur et la grille de distribution. Pour ce faire, il est nécessaire de mettre en place un système de communication de haut niveau entre le contrôleur de la charge du véhicule (Electric Vehicle Communication Controller, ou EVCC) et un contrôleur de borne intelligent (Supply Equipment Communication Controller, ou SECC). Cette possibilité permet aussi d'authentifier automatiquement un utilisateur qui a un contrat préalablement établi avec l'opérateur de borne ou de mobilité et lui fournir le service à valeur ajoutée comme le partage de la connexion internet au VE par une borne de recharge**. [7]**

La communication véhicule-borne avec l'ISO/IEC 15118 explique simplement un protocole de communication qui fonctionne au-dessus d'une communication TCP-IP.

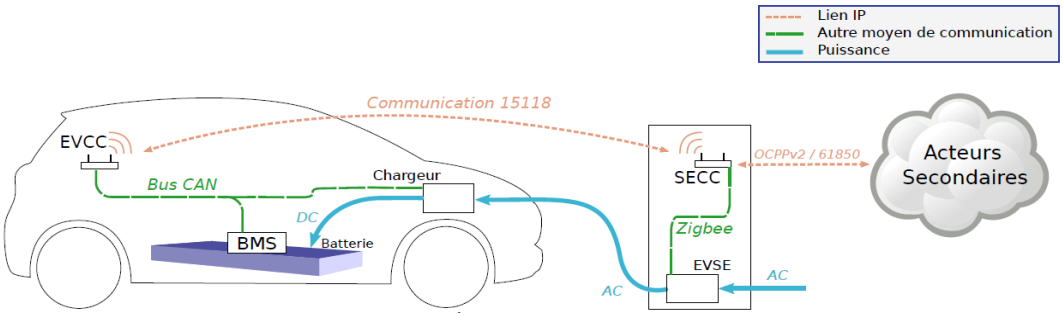

Figure 2.4. : Implantation possible d'un système de charge ISO/IEC 15118 dans le cas de la charge AC [7]

Ce protocole de communication se base seulement sur le protocole IPv6 avec une bonne utilisation de la couche routable, il permet une bonne flexibilité dans l'architecture des réseaux. Il est aussi exploité dans la communication de tous les éléments de la charge de véhicule électrique.

Enfin, la possibilité de communication ~~de~~ bout en bout **[11]** permet de ne pas fixer à l'avance le placement des entités communicantes, qui pourra changer au fur et à mesure de l'évolution du modèle de chargement des VE. La communication définie par le protocole ISO/IEC 15118-2 est Stateful (à états), c'est-à-dire que les messages échangés entre le véhicule et la borne dépendent d'états propres à chaque entité.

Couche	Protocole	Description
7	Messages applicatifs	Messages applicatifs tels que décrits dans la norme.
6	EXI	Efficient XML Interchange : format XML binaire, pour réduire la taille des messages et la durée d'analyse.
5	V2GTP	En-tête de communication V2G (version du protocole, type d'encodage du message, taille du payload).
4	TCP, UDP, TLS	TCP : Protocole utilisé pour les échanges ISO/IEC 15118-2, façon client (EVCC) / serveur (SECC). UDP : Utilisé pour la découverte du SECC. TLS : Permet de sécuriser les échanges TCP (obligatoire dans la plupart des cas d'usage).
3	IPv6	Protocole réseau utilisé pour les échanges.
2	PLC / IEEE	Communication sur un lien PLC (ISO/IEC 15118-3).
1	802.11x	Communication sur un lien WiFi (ISO/IEC 15118-7, en cours de rédaction).

Tableau 2.1. Modèle OSI du protocole définit par ISO/IEC 15118 [7]

2.5. Conclusion

Le réseau véhiculaire V2G étant un système, plusieurs ressources doivent intervenir dans son fonctionnement et chacune d'elle joue un rôle précis dans le but ~~est~~ de recharger un VE avec moins de~~s~~ risques de nuisance.

Pour permettre une bonne interopérabilité des ressources et une bonne communication, la norme ISO 15118 avec ses parties a été créée pour définir les exigences, spécifications, les lignes de base de ce système. Elle assure aussi une meilleure prévention contre la cybercriminalité et prévoit une bonne protection des données des conducteurs. **[8]**

Si la première partie de la norme ISO 15118 éditée en 2013 spécifie et définit les termes attachés au réseau V2G puis décrit les exigences générales et les cas d'utilisation, elle donne aussi une présentation générale et une compréhension commune des aspects influant sur le processus de la charge, de la facturation et du paiement. L'ISO 15118-1 ne spécifie pas la communication interne du véhicule entre la batterie et l'équipement de charge, ni la communication du SECC à d'autres acteurs et à l'équipement.

La deuxième partie de l'ISO 15118 cependant spécifie seulement la communication entre les véhicules électriques à batterie (BEV) ou les véhicules électriques hybrides rechargeables (PHEV) et l'équipement de distribution de véhicule électrique.

Le groupe de messages de la couche application définie dans la deuxième partie de l'ISO 15118 est conçu pour prendre en charge le transfert d'énergie d'un EVSE à un EV, les messages, le modèle de données, le format de représentation des données XML / EXI, l'utilisation de V2GTP, TLS, TCP et IPv6. En outre, il décrit comment accéder aux services de la couche liaison de données.

Le **Tableau 2.2** ci-après présente les parties de la norme ISO15118 et les dates de publication :

Document	**Titre**	**Statut**
ISO 15118-1	Informations générales et définition des cas d'usage	Publiée en 04/2013
ISO 15118-1 Ed.2	Informations générales et définition des cas d'usage associés à l'ISO 15118-20	Publiée en 04/2019
ISO 15118-2	Exigences des protocoles réseau et application	Publiée en 03/2014
ISO 15118-3	Exigences des couches physiques CPL et accès aux données	Publiée en 05/2015
ISO 15118-4	Tests de conformité des protocoles réseau et application	Publiée en 05/2018
ISO 15118-5	Tests de conformité de la couche physique CPL et d'accès aux données	Publiée en 05/2018
ISO 15118-8	Exigences des couches physiques et accès aux données pour la communication sans fil	Publiée en 05/2018
ISO 15118-9	Tests de conformité des couches physique Wifi et d'accès aux données	En rédaction (Comitee Draft), publication prévue en 2021
ISO 15118-20	Exigences des protocoles réseau et application de deuxième génération	2020.

Tableau 2.2. : Parties de la norme ISO15118 [8]

Il ressort après l'étude du réseau V2G que la sécurité occupe une place de choix. C'est pour cette raison que dans le chapitre suivant, nous présenterons une étude des systèmes de détection d'intrusion afin de comprendre ces concepts.

CHAPITRE 3 : SYSTÈMES DE DÉTECTION D'INTRUSIONS

3.1. Introduction

Un système (surtout informatique) où plusieurs objets sont autorisés à accéder est toujours exposé aux attaques, car il peut toujours exister une certaine vulnérabilité. Cette vulnérabilité peut entrainer un accès des objets indésirables et non autorisés au système. D'où la nécessité pour les gestionnaires des systèmes informatiques de toujours avoir un système de détection et de prévention contre les intrusions.

Dans un réseau de communication, il y a plusieurs outils de sécurité et les techniques de détection des anomalies sont la dernière ligne de défense lorsque d'autres approches ne parviennent pas à détecter les menaces de sécurité ou autres problèmes. **[12]**

Afin d'optimiser le système de détection d'intrusions dans les réseaux véhiculaires V2G, nous allons d'abord dans ce chapitre présenter une étude sous forme tutorial sur les attaques, les IDS (Intrusion Detection System) et les IPS (Intrusion Prevention System).

3.2. Définition et présentation d'un IDS

Le système de détection d'intrusions consiste à découvrir ou identifier l'utilisation d'un système informatique à d'autres fins que celles prévues. **[13]**

C'est grâce aux différents systèmes de détection d'intrusion (IDS) que les organisations, entreprises et autres structures peuvent protéger leur système contre les menaces venant de l'intérieur ou de l'extérieur d'un réseau d'objets.

Tous les utilisateurs qui se connectent sur un réseau public courent toujours des risques. Ces risques sont à la base de plusieurs dommages dans le système, ce qui risque de déstabiliser le bon fonctionnement de celui-ci. Les hackers malicieux~~ses~~ abusent toujours des vulnérabilités des services, des applications ou du réseau pour s'attaquer à un terminal ou à un système. C'est donc pour avoir une bonne sécurité que les IDS sont toujours importantes dans un système. C'est le cas du réseau véhiculaire V2G (Vehicle-To-Grid) qui ne doit pas admettre des intrusions lors des échanges des données entre le VE (Electric Vehicle) et la borne.

Les IDS jouent donc un rôle comparable à celui d'une caméra de surveillance installée devant le port pour surveiller toutes les attaques destinées au réseau. **[14]**

3.3. Description d'un système de détection des intrusions (IDS)

Un système de détection des intrusions est une combinaison ou intégration de plusieurs services pour assurer une meilleure gestion et contrôle de sécurité du système.

Dans la **Figure 3.1**, nous illustrons un exemple de modèle d'un IDS proposé par le groupe IDWG (Intrusion Detection exchange format Working Group) **[15].**

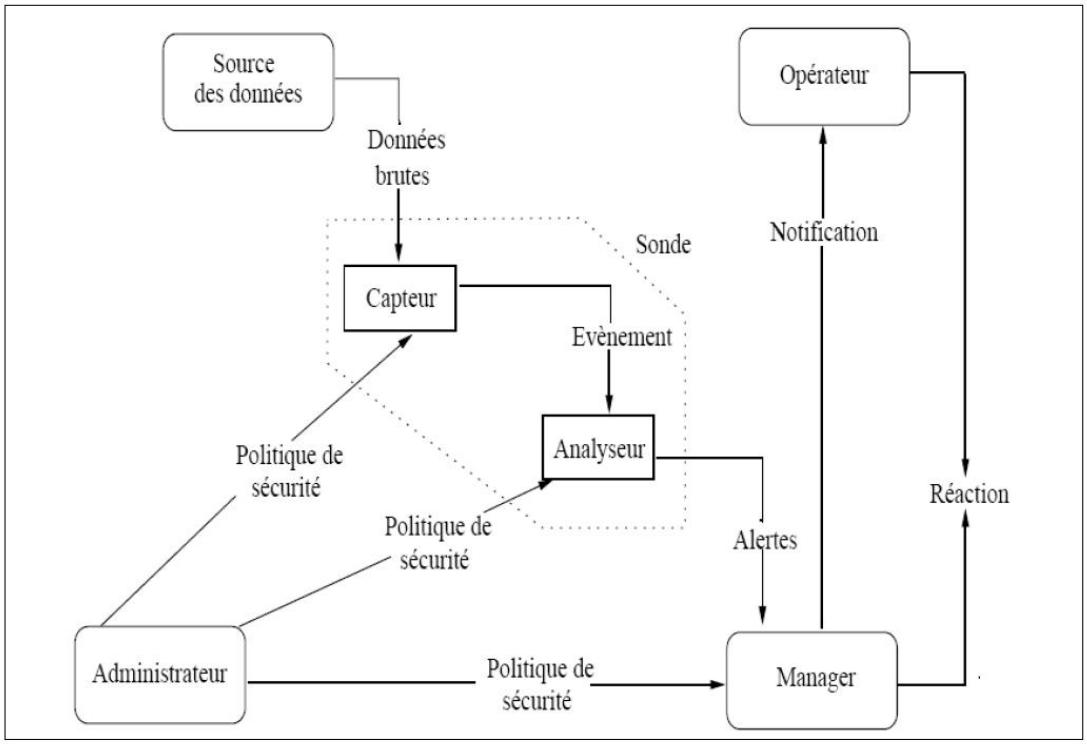

Figure 3.1. : Modèle général d'un IDS proposé par le groupe IDWG (Intrusion Detection exchange format Working Group) [15]

Où :

- **Alerte** : est un message formaté et émis par un analyseur lorsqu'il y a des activités intrusives contre une source de données.
- **Analyseur** : c'est un outil matériel ou logiciel qui met en œuvre l'approche choisie pour la détection (comportementale ou par scénarios), il génère des alertes lorsqu'il détecte une intrusion.
- **Capteur** : un logiciel générant des événements en filtrant et en formatant les données brutes provenant d'une source de données.

- **Événement** : un message formaté et envoyé par un capteur. C'est l'unité élémentaire utilisée pour représenter un élément d'un scénario d'attaque.
- **Manager** : composant d'un IDS permettant à l'opérateur de configurer les différents éléments d'une sonde et de gérer les alertes reçues et éventuellement la réaction de l'opérateur.
- **Notification** : la méthode par laquelle le manager d'un IDS met au courant l'opérateur de l'occurrence d'une alerte.
- **Opérateur** : une personne chargée de l'utilisation du manager associé à l'IDS, elle décide de la réaction à apporter en cas d'alerte, elle est parfois la même entité que l'administrateur.
- **Politique de sécurité** : c'est la spécification des exigences de sécurité à faire respecter dans un réseau d'une organisation afin de garantir l'intégrité, la confidentialité et la disponibilité des ressources sensibles. Elle définit quelles activités sont autorisées et lesquelles sont interdites.
- **Réaction** : sont les mesures passives ou actives prises en réponse à la détection d'une attaque, dans le but de la stopper ou pour corriger ses effets.
- **Sonde** : un ou plusieurs capteurs couplés avec un analyseur.
- **Source de données** : dispositif générant de l'information sur les activités des entités du système à surveiller. **[15]**

3.4. Attaques

En informatique, une attaque est une forme d'exploitation d'une faille d'un système informatique à des fins non connus par l'exploitant du système et est généralement préjudiciable. [**16**]

Un pirate informatique qui attaque un système peut le faire pour divers buts tels que de s'emparer des informations sensibles du réseau ou pour des raisons purement commerciales, pour terroriser l'entreprise, espionner le réseau, attirer l'attention, avantages concurrentiels, vérification de la sécurité d'un système. **[16]**

3.4.1. Étapes d'une attaque

Pour réaliser une attaque, plus étapes doivent être respectées dont :

- **La reconnaissance des informations** concernant la victime qu'il veut attaquer avant même d'utiliser les outils appropries de l'attaque;
- **Gain d'accès (Gain Access)** qui permet au pirate d'avoir un accès aux ressources des victimes**.** Le niveau d'accès requis dépend évidemment de l'attaque. Notons toutefois que

certains types d'attaques, comme les attaques en déni de service, n'ont pas besoin d'accès sur la machine victime;

- **Augmentation de privilèges (Privilege Escalation) :** dans le but de connaitre les comptes et autres ressources qui ont plus de priorité, de certificats … afin de bien planifier son attaque;
- **Actions principales (Principal Actions) :** cette étape peut prendre différentes formes par exemple, l'attaquant peut exécuter une attaque en déni de service, installer un code malveillant, compromettre l'intégrité des données ou exécuter un programme.
- **Cacher les traces (Hiding Traces) :** les attaquants les plus expérimentés utilisent généralement cette dernière étape pour effacer leurs traces et rendre ainsi la détection plus difficile. **[17]**

3.4.2. Types d'attaques

Une attaque peut être :

7. **Directe :** C'est-à-dire que l'attaquant attaque directement la victime via son terminal comme le montre la **Figure 3.2**. **[16]**

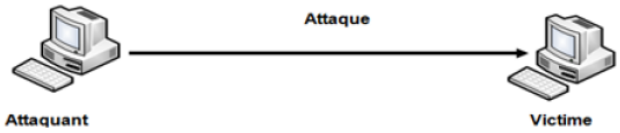

Figure 3.2. : Attaque directe

8. **Indirectes par rebond** ou l'attaquant s'infiltre dans le réseau et attaque la victime par un terminal intermédiaire comme le montre la **Figure 3.3.** Dans ce cas, il est très difficile de connaitre l'identité de l'intru car l'attaquant ne dévoile pas ses ressources.

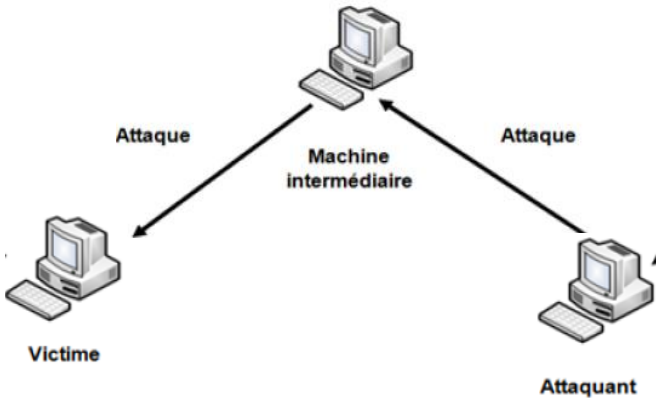

Figure 3.3. : Attaque indirecte par rebond

9. **Indirectes par réponse** qui sont des attaques dérivées des celles par rebond. Elles offrent les mêmes avantages, du point de vue de l'intrus. Mais au lieu qu'il envoie une attaque à la

machine, il lui envoie une requête et la réponse à cette requête sera envoyée à la victime. **[16]**

3.4.3. Techniques d'attaque informatique

En informatique, tous les terminaux qui sont connectés à un réseau privé ou public sont potentiellement vulnérables aux attaques : c'est le cas des VE (Véhicule Électrique) dans les réseaux V2G. Ces attaques sont perpétrées par des pirates qui se servent de plusieurs techniques selon les objectifs qu'ils veulent atteindre ou selon leurs compétences techniques.

Dans ce paragraphe, nous allons résumer quelques techniques d'attaques les plus usuelles.

3.4.3.1. Attaques par déni de service (DoS) et par déni de service distribué (DDoS)

Une attaque DOS ou DDOS concerne les ressources d'un système informatique dans le but de les empêcher à fonctionner normalement ou de mettre hors service (hors ligne) un système afin de permettre à l'attaquant de lancer une autre attaque. **[18]** Quand elle part de plusieurs sources, on parle de DDOS et ces attaques peuvent concerner les services d'un routeur, d'un serveur **[19]** que l'attaquant est capable de bloquer et prendre sa possession puis le rendre indisponible en perturbant les connexions entre les terminaux et l'accès au service.

Parmi les attaques les plus courants de DOS et DDOS, on retrouve:

Attaque TCP SYN Flood : cette attaque est capable de planter le système jusqu'à le rendre inutilisable. Comme les segments SYN (synchronization) sont utilisés par le protocole TCP (Transmission Control Protocol) lors de l'établissement de la connexion entre le client et le serveur, cette attaque permet d'envoyer plusieurs requêtes successivement de SYN vers la victime sans qu'il ne reçoive des ACK (Acknowledge).

Attaque teardrop : Cette attaque provoque le chevauchement des champs de longueur et de décalage de fragmentation des paquets séquentiels du protocole Internet (IP) au niveau de l'hôte attaqué ; au cours de ce processus, le système attaqué tente de reconstruire les paquets, mais échoue. Le système cible embrouille et plante. **[18]**

Attaque Smurf : Ce type d'attaque DOS utilise la technique du "Broadcast Ping" afin que le nombre de paquets ICMP envoyés à la station grandisse de manière exponentielle causant alors un crash presque inévitable. **[19].** Elle tente d'anticiper l'adresse IP et utilise les messages ICMP (Interchange Message Message Protocol) pour saturer le trafic du système cible. **[18]**

Ping of death : c'est un type d'attaque qui consiste à envoyer un mauvais ping à la cible. Ce ping a une déformation, car les paquets IP (Internet Protocol) qui sont utilisés dans l'adresse ont une taille plus élevée que le maximum recommandé qui est de 65.535 bytes. Comme les paquets IP qui sont supérieurs à la taille maximum ne peuvent pas être transmis dans le réseau, l'attaquant les fragmente~~nt~~ et à la réception par la cible, ils seront réassemblés et vont provoquer un débordement de la mémoire tampon qui a son tour peut provoquer une panne de dysfonctionnement du service.

Botnets : Ces types d'attaques s'appuient plus sur le vol des données, l'envoi des spam, l'accès au terminal cible et de sa connexion. Les botnets sont des réseaux constitués de millions de systèmes infectés par des logiciels malveillants et contrôlés par des pirates informatiques afin d'effectuer des attaques DDoS. **[18]**

3.4.3.2. Attaque de l'homme au milieu (MitM : Man in the Middle)

Une attaque de l'homme du milieu est un pirate qui s'insère dans les communications entre un client et un serveur **[18].** Il peut faire :

Le Détournement de session : (en anglais « session hijacking ») c'est un type d'attaque qui consiste à détourner les données qui étaient destinées à un terminal cible pour les rediriger vers le pirate attaquant. À ce moment, le serveur continu a communiqué comme s'il communiquait avec un client alors qu'il échange avec un pirate.

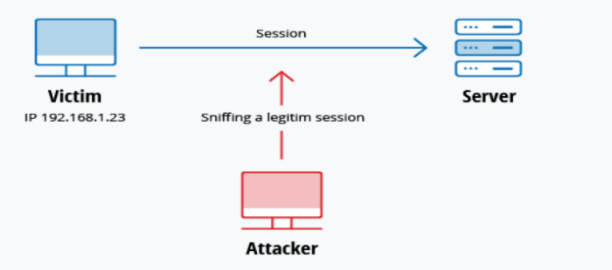

Figure 3.4. : Attaque par détournement de session (phase 1) [27]

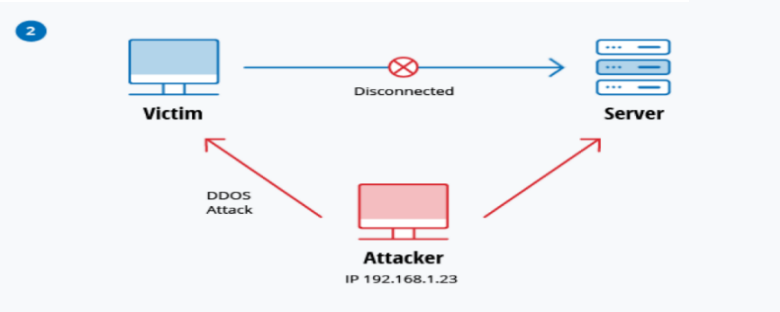

Figure 3.5. : Attaque par détournement de session (phase 2) [18]

Usurpation d'IP : Un pirate peut utiliser l'usurpation d'adresse IP pour convaincre un système qu'il communique avec une entité connue et fiable afin de lui donner accès au système. Le pirate envoie à un hôte cible un paquet contenant l'adresse IP source d'un hôte connu et fiable au lieu de sa propre adresse IP source. Il est possible que l'hôte cible accepte le paquet et agisse en conséquence. **[18]**
Relecture : appelée en anglais : replay attack ou playback attack, c'est une forme d'usurpation d'identité qui se produit par répétition malicieuse d'un pirate qui a intercepté la transmission des données qui ne lui étaient pas destinées.

3.4.3.3. Attaques phishing et spear phishing

Appelée en français hameçonnage, c'est une technique qui permet au pirate d'obtenir des renseignements personnels de la victime dans le but de perpétrer une usurpation d'identité en envoyant des courriels qui semblent provenir de sources fiables dans le but d'obtenir des informations personnelles ou d'inciter les utilisateurs à faire quelque chose. Cette technique combine ingénierie sociale et stratagème technique. **[18]**

3.4.3.4. Attaque par Drive by Download

Appelée en français téléchargement furtif ou téléchargement à la dérobée, c'est une technique d'attaque qui traduit un téléchargement involontaire des logiciels (espion, virus, malveillants, criminels) sur internet. Ce téléchargement de produit parfois à l'insu de l'utilisateur lors de l'ouverture d'une pièce jointe, d'un lien ou de la visite d'un site web non sécurisé en insérant un script malveillant dans le code HTTP ou PHP de l'une des pages. Ce script est capable d'installer des logiciels malveillants directement dans le terminal de la victime lors de la visite du site ou le rediriger vers un autre site web contrôlé par les attaquants. **[18]**

3.4.3.5. Attaque par mot de passe

Plusieurs systèmes informatiques utilisent couramment le mot de passe comme mécanisme de sécurité avant d'y accéder. Il est considéré comme un important authentificateur des utilisateurs autorisés et c'est pour cette raison que les pirates l'exploitent pour attaquer le système en espérant avoir un accès dans son compte. Le mot de passe de la victime peut être obtenu en fouillant son bureau physique, en surveillant sa connexion au réseau pour acquérir des mots de passe non chiffrés, en ayant recours à l'ingénierie sociale, en accédant à une base de données de mots de passe ou simplement en devinant. **[18]**

3.4.3.6. Attaque par écoute illicite

Il fait partie des types d'attaques appelées « écoutes clandestines » qui ne sont autres que le résultat d'une interception du trafic réseau par un attaquant. Ce type d'attaque permet à un attaquant d'obtenir des mots de passe, des numéros de carte bancaire et d'autres informations confidentielles qu'un utilisateur envoie sur le réseau. Elles peuvent être passives ou actives.

Une écoute clandestine est dite passive lorsque le pirate détecte des informations en écoutant la transmission de messages sur le réseau. Elle est dite écoute clandestine active lorsque le pirate s'empare activement d'informations en se faisant passer pour une unité amie et en envoyant des requêtes aux transmetteurs. On appelle cela sonder, scanner ou saboter. **[18]**

3.4.3.7. Attaque par des logiciels malveillants

Appelé en anglais attaque malware, elle se produit généralement lorsqu'un pirate développe un logiciel malveillant qui est ensuite installé dans le terminal de la victime à son insu et sans son consentement dans le but de permettre au pirate d'avoir un accès aux données de l'utilisateur ou d'endommager son terminal. Les logiciels malveillants les plus courants sont : Macro-virus, Infecteurs de chiers, Infecteurs de système ou de secteur d'amorçage, Virus polymorphes, Virus furtifs, Chevaux de Troie, Bombe logique, Vers, Injecteurs, Rançongiciels (ransomware), Logiciels publicitaires (adware), Logiciels espions (spyware).

3.4.4. Analyse des attaques

L'analyse du trafic peut différer d'une architecture à une autre, nous pouvons distinguer deux types d'analyses : l'analyse locale centralisée et l'analyse distribuée. L'analyse peut aussi être périodique ou continue. Pour analyser un trafic, nous pouvons distinguer quatre tâches différentes :

Agrégation des données : permets de faire la collecte d'informations et la normalisation du format des données pour faciliter le traitement.

Réduction des données : permets de filtrer les données inutiles, trouver les redondances et éliminer les données erronées.

Corrélation des données : permets d'identifier les relations entre les alertes dans le but de les regrouper (Clustering) en fonction de différentes variables (temps, évènement, port destination, protocole, contenu du message, etc.).

Induction des données : essayer de comprendre de nouvelles données, découvrir de nouveaux patrons ou identifier des nouvelles attaques. **[15]**

	Vrai	*Faux*
Positif	Une alerte est générée et une condition présente doit être révélée	Une alerte et générée et aucune condition présente n'est révélée
Négatif	Une alerte n'est pas générée et aucune condition présente n'est pas révélée	Une alerte n'est pas générée et une condition présente doit est révélée

Tableau 3.1. Alertes Vrai/Faux et Positif/Négatif [15]

3.5. Réaction et comportement après une attaque

Comme nous l'avons expliqué précédemment, l'administrateur du système est la ressource principale qui définit la politique de sécurité d'un système informatique, donc même la qualité et la réaction d'un IDS dépend de la politique qu'il a mis en place. Généralement, en cas d'une attaque, un IDS peut réagir selon une approche active ou passive.

Avec **l'approche active**, l'IDS au-delà de toutes les réactions possibles face à une attaque (réinitialisation de la connexion, blocage du trafic, suppression de tous les processus du système attaquant, etc.), envoie des alertes soit en mode complet en donnant tous les détails possibles ou soit en mode réduit qui consiste à émettre une alerte avec des informations essentielles seulement (Full Mode Alert, Fast Mode Alert). **[15]**

Avec l'**approche passive,** l'IDS n'émet aucune alerte dans le système, mais il envoie juste un courriel ou un message court (SMS) à l'administrateur du système. **[15]**

3.6. Classification des IDS

Les IDS sont classés en 3 catégories **[20]** :

- **IDS systèmes ou IDS hôtes (HIDS)** : sont basés sur un hôte qui peut être un poste de travail, un serveur ou autres terminaux. Appelé en anglais HIDS (Host-based IDS), ils analysent seulement les données qui concernent cet hôte. Par manque de contrôleur de trafic, ils utilisent deux types de sources pour fournir une information sur l'activité : les logs (journaux) et les traces d'audit du système d'exploitation. **[20]**
- **IDS réseaux (NIDS)** : appelés en anglais NIDS (Network-based Intrusion Detection System), ils analysent et interprètent les paquets des données qui circulent dans ce réseau. Ils fabriquent des signatures et utilisent des détecteurs pour analyser les trafics et programmer l'envoi d'une alerte afin de bien identifier les intrus.

L'IDS peut utiliser les couches réseau (IP : Internet Protocol, ICMP : Interchange Message Message Protocol), transport (TCP : Transmission Control Protocol, UDP : User Datagram Protocol)

et application (TELNET : Telecommunication Network, HTTP : Hypertext Transfer Protocol , SNMP : Simple Network Management Protocol,) pour accomplir sa fonction, il est capable de détecter des paquets malveillants conçus pour contourner un coupe-feu aux règles de filtrage trop laxistes (indulgentes), et de chercher des signes d'attaque à différents endroits sur le réseau. **[20]**

- **IDS hybrides** : ce sont des IDS qui sont constituées par une association des deux IDS : HIDS et de NIDS.

On peut aussi parler d'autres familles d'IDS comme :

- Les IDS de nœud réseau en anglais (NNIDS pour Network Node IDS) fonctionnent comme les NIDS classiques,
- Les IDS basés sur une application (ABIDS) pour Application-Based IDS, c'est un sous-groupe des IDS hôtes **[20].**

3.7. Composants d'un IDS

Étant donné que l'analyse du trafic d'un réseau est la principale fonction d'un IDS, tous les trafics qui y circulent sont considérés comme des potentiels malveillants. C'est pourquoi il faut protéger les systèmes par des IDS qui ne doivent pas être désactivées et/ou corrompus **[20**].

Afin de renforcer son efficacité, un IDS fonction avec d'autres composants comme par exemple, les sondes, la console de gestion, le concentrateur d'évènements, la console d'alerte pour informer le système en cas de menace.

3.7.1. Sonde

C'est un équipement passif qui est chargé spécialement d'écouter le flux des données d'un réseau puis de faire remonter les informations et les alertes en fonction de son point d'installation. Il localise les signaux générés par une cyberattaque et qui sont faibles en faisant passer librement l'ensemble des flux de données. **[21]**

3.7.2. Console de gestion

C'est une sorte de conteneur qui contient les outils de la gestion de l'IDS, les interfaces de configuration et autres fonctions nécessaires au fonctionnement de l'IDS.

3.7.3. Concentrateur d'évènements

C'est un outil de l'IDS qui est doté d'une haute disponibilité et qui est chargé de recevoir les messages qui viennent des sondes. Les messages reçus peuvent être stockés, relayés ou filtrés.

3.7.4. Console d'alerte pour informer le système en cas de menace

C'est un outil de l'IDS qui est chargé de gérer des alertes, les conditions qu'il faut pour émettre une alerte, d'afficher les informations sur les alertes …

3.8. Techniques ou approches de détection d'intrusion

Il existe deux techniques de détection d'intrusion qui sont généralement utilisées de façon complémentaire :

- **La détection d'abus (misuse detection) ou par signature** : on l'appelle aussi détection de mauvaise utilisation ou malveillance. Elle consiste à analyser l'information reçue, la compare avec une base de données de signatures d'attaques, pour voir si s'est considérée comme une attaque.
- **La détection d'anomalies (anomaly detection)** : Elle est plus ancienne et permet de détecter les comportements suspects en téléphonie. Elle permet de modéliser le comportement "normal" d'un système, programme, utilisateur pendant une certaine période en définissant une ligne de conduite appelée vaseline ou profil, et de considérer ensuite (en phase de détection) comme suspect tout comportement inhabituel (les déviations significatives par rapport au modèle de comportement "normal). **[20]**

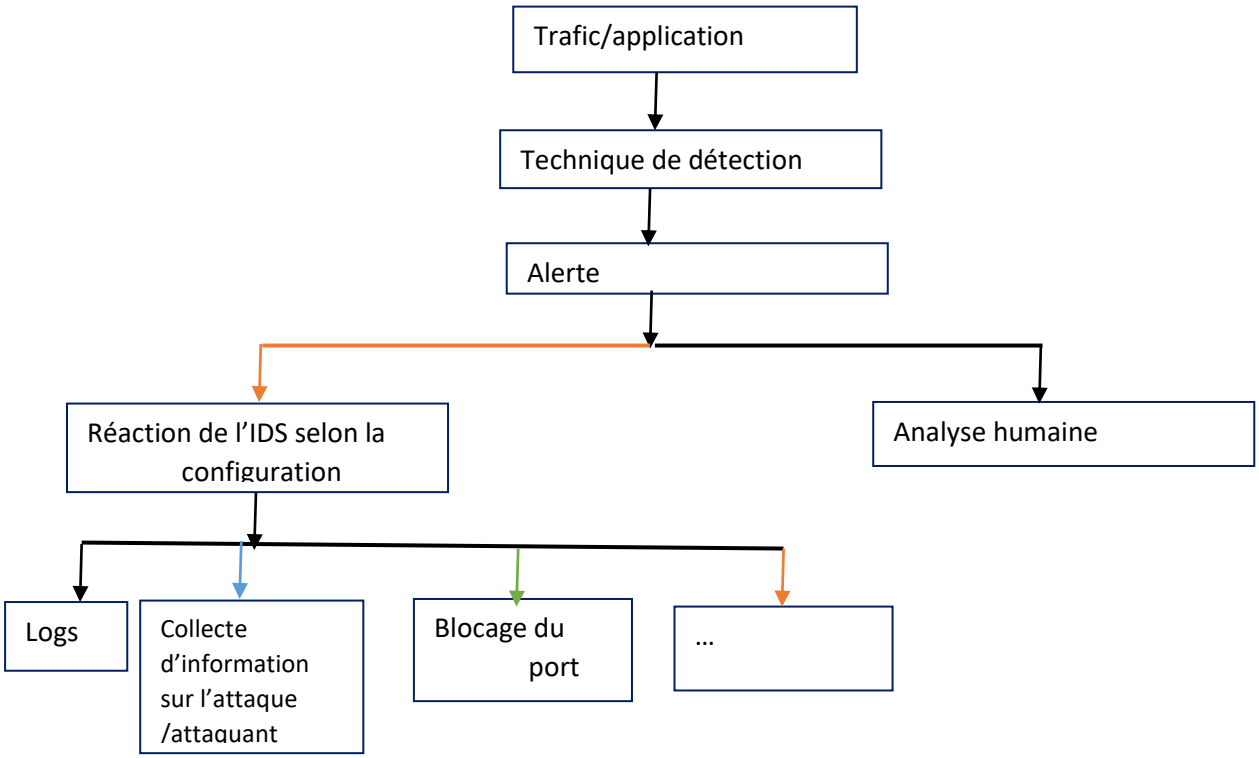

Figure 3.6. Fonctionnement d'un IDS [20]

Les IDS peuvent prendre 3 états : fonctionnement (WST), protégé (Pst) et dangereux (Dst). **[22]**

3.9. Limites des IDS

Les IDS présentent aussi des limites qui sont remarquables dans les deux techniques de détection et peuvent se montrer impuissantes dans certains cas comme :

3.9.1. Attaques sur les drapeaux TCP

- Envoi d'un faux SYN (synchronization) ;
- Intégration de données avec un mauvais numéro de séquence ;
- FIN/RST spoofing avec mauvais numéro de séquence ;
- Désynchronisation après connexion ;
- Désynchronisation avant connexion [SYN (mauvaise somme de contrôle + mauvais numéro de séquence) puis SYN] ;
- FIN/RST spoofing avec mauvaise somme de contrôle ;
- Data spoofing avec mauvaise somme de contrôle ;
- FIN/RST spoofing avec TTL (Time-to-live) court.

- Insertion de données avec un TTL court, etc. **[20]**

3.9.2. Placement de l'IDS

Qui concerne plus la détection d'intrusion dans la zone démilitarisée, dans le(s) réseau(x) privé(s) et derrière le pare-feu (détection des signes parmi tout le trafic entrant et sortant).

3.9.3. Pollution/surcharge

C'est une faiblesse qui est dû à la surcharge des alertes des IDS qui saturent ses ressources comme la mémoire, disque dur, processeur, occasionnant parfois la perte de paquets, le déni de service partiel ou total… **[20]**

3.9.4. Contournement/évasion

Les IDS peuvent également être contournés par un rejet des paquets corrects. Cela peut être causé par une mauvaise définition et maintenance des signatures, de la mise à jour de la base, la charge du système qui viennent des aspects ci-après :

- Limites "humaines" : signatures pas à jour ou mal conçues
- Mauvais contexte d'utilisation des signatures
- Vulnérabilité aux mutations des signatures
- Les faux positifs dus aux mauvais alertes **[20]**

3.10. Prévention d'intrusion

La prévention d'intrusion est un ensemble de technologies de sécurité ayant pour but d'anticiper et de stopper les attaques. La prévention d'intrusion est appliquée par quelques IDS récents et diffère des techniques de détection d'intrusion décrites précédemment. Ainsi, au lieu d'analyser les logs du trafic, c'est-à-dire découvrir les attaques après qu'elles se soient déroulées, la prévention d'intrusion essaie de prévenir ces attaques.

Là où les systèmes de détection d'intrusion se contentent de donner l'alerte, les systèmes de prévention d'intrusion bloquent le trafic jugé dangereux. **[20]**

3.10.1. Systèmes de prévention d'intrusion (IPS)

Les IPS et les IDS présentent un système de fonctionnement semblable avec les mêmes types (IPS hôte et IPS réseau), sauf que les IPS sont capables d'effectuer l'analyse des contextes de connexion, des logs et la coupure des connexions suspectes. Contrairement aux IDS surtout classiques, ceux-ci peuvent fonctionner sans une surveillance constante, car ils stoppent les attaques du réseau et n'utilisent pas de signature pour détecter une attaque. Ils ont les fonctions suivantes :

- Surveillance du comportement d'application,
- Création de règles d'application,
- Alerte causée par des violations,
- Corrélation avec d'autres événements,
- Interception d'appels au système **[23]**

3.10.2. Classification des IPS

Les systèmes de prévention des intrusions peuvent être classés en quatre types différents **[24] :**

1. **Système de prévention des intrusions en réseau (NIPS)** : surveille l'ensemble du réseau à la recherche de trafic suspect en analysant l'activité du protocole. **[24]**
2. **Système de prévention des intrusions sans fil (WIPS)** : surveille un réseau sans fil pour détecter tout trafic suspect en analysant les protocoles du réseau sans fil. **[24]**
3. **Analyse du comportement du réseau (NBA)** : examine le trafic réseau pour identifier les menaces qui génèrent des flux de trafic inhabituels, tels que les attaques par déni de service distribué (DDoS), certaines formes de logiciels malveillants et les violations de règles. **[24]**
4. **Système de prévention des intrusions basé sur l'hôte (HIPS)** : un logiciel installé qui surveille les activités suspectes d'un seul hôte en analysant les événements qui se produisent sur cet hôte. Il est capable de reporter des alertes ou d'y réagir lui-même. **[24]**

3.10.3. Méthodes de détection des IPS

La majorité des systèmes de prévention des intrusions utilisent l'une des trois méthodes de détection suivantes : **[24]**

1. **Détection basée sur les signatures** : L'IDS basé sur les signatures surveille les paquets dans le réseau et compare avec les modèles d'attaque préconfigurés et prédéterminés connus sous le nom de signatures. **[24]**
2. **Détection statistique basée sur les anomalies** : Un IDS basé sur les anomalies surveillera le trafic réseau et le comparera à une base de référence établie. La ligne de base identifiera ce qui

est “normal” pour ce réseau, quel type de bande passante est généralement utilisé et quels protocoles sont utilisés. Elle peut cependant déclencher une alarme de faux positif pour une utilisation légitime de la bande passante si les lignes de base ne sont pas configurées intelligemment. **[24]**

3. **Détection d’analyse de protocole dynamique** : Cette méthode identifie les déviations des états du protocole en comparant les événements observés avec des profils prédéterminés de définitions généralement acceptées de l’activité bénigne. **[24]**

3.10.4. Limites des IPS

Les limites qu’il faut signaler pour les IPS restent la délicatesse de sa mise en place, leur administration complexe ce qui peut bloquer tout le réseau en cas de fausse alerte et enfin le manque de standard.

3.11. Conclusion

En conclusion de cette présentation sur les IDS, nous avons constaté qu'en plus des IDS, il faut avoir des IPS qui sont également très efficaces et qui complètent la protection d'un système.

Il faut aussi reconnaitre que l’efficacité des IDS dépend de la configuration que l’administrateur du système va leur donner, car c’est lui qui définit les spécifications d’attaques et de défense de son IDS. Elle dépend aussi de sa robustesse (résistance aux défaillances) et de la faible quantité de faux positifs (fausses alertes) et de faux négatifs (attaques non détectées) qu'il génère.

Les IDS et les IPS sont bien des outils techniques qui ne manquent pas d'inconvénients et de limites et les humains sont obligés d’intervenir à chaque instant afin d’améliorer leurs performances.

Ce sont des solutions qui offrent une bonne sécurité pour un système, ou un réseau de communication mais, ils ne garantissent pas une sécurité à 100% dans la mesure où pendant que ces systèmes connaissent des améliorations, les systèmes d’attaque aussi sont améliorés dans le but de percer les frontières établies sans qu’ils ne génèrent des alertes à temps réel. C’est pourquoi, il est important de chercher à améliorer les IDS avec des méthodes ou modèles appropriés afin de les rendre beaucoup plus efficaces.

Dans le chapitre qui suit, nous allons faire une étude descriptive des méthodologies (règles d’association et la régression logistique) que nous allons utiliser pour améliorer les IDS.

CHAPITRE 4 : RÈGLES D'ASSOCIATION MAXIMALE ET LA RÉGRESSION LOGISTIQUE

4.1. Introduction

Les composants qui sont utilisés dans la sécurité des réseaux présentent toujours une vulnérabilité, voilà pourquoi, à côté de ces matériels sont souvent associés des IDS (Intrusion Detection System) qui constituent un autre mur de protection permettant de protéger les systèmes et d'identifier les intrusions.

Plusieurs solutions existent pour améliorer les IDS et nous avons porté notre choix sur les règles d'association maximale qui seront utilisées pour améliorer significativement les IDS dans les réseaux véhiculaires V2G (Vehicle-To-Grid) et plus précisément entre le VE (Electric Vehicle) et la borne de recharge.

Les règles d'association maximale seront complétées par une méthode statistique qui est la régression logistique multivariée.

Dans ce chapitre, nous allons aborder notre état de l'art avec la description des règles d'association et puis de la régression logistique.

4.2. Règles d'association

Les règles d'association étaient plus utilisées dans le domaine du data mining pour mener une étude profonde sur la découverte des relations entre les variables stockées dans une importante base de données. Elles sont aujourd'hui utilisées dans plusieurs domaines comme celui de la fouille du web, de la détection d'intrusions et de la bio-informatique. Elles sont de types A implique B. **[25]**

La plupart des algorithmes utilisés dans ce domaine, parcourent les données pour trouver les éléments qui dépassent un support minimum défini par l'utilisateur, et extraient par la suite, les règles d'association dont la confiance dépasse une confiance minimum.

On peut bien mesurer la force d'une règle d'association à partir de son support et de sa confiance que nous allons expliquer dans les lignes qui suivent. **[25]**

4.2.1. Concepts

- **Transaction :** c'est un sous-ensemble qui est inclus dans un ensemble ;
- **Règles d'association :** c'est l'évaluation d'une relation qui se base principalement sur le calcul des paramètres appelés support et confiance. Elles aident à montrer la probabilité de relations entre les éléments de données au sein de grands ensembles de données dans divers types de bases de données. **[26] [27].** Elles se présentent sous la forme X→Y (règle d'association entre A et B) ou X et Y sont des ensembles d'items disjoints ; **[25]**
- **Itemset:** C'est un ensemble d'objets ou d'articles ;
- **Support d'une règle d'association:** c'est le nombre de transactions qui contiennent une association (A et B par exemple) divisée par le nombre total des transactions. Il permet de fixer un seuil en dessous duquel les règles ne sont pas considérées comme fiables; **[26]**
- **Confiance** : elle représente la précision de la règle et peut être vue comme la probabilité conditionnelle ; **[26]**

Hacène Cherfi et Yannick Toussaint précisent que la confiance mesure le degré de validité d'une règle, c'est-à-dire lorsqu'il existe des contre-exemples de documents qui vérifient B, mais pas nécessairement tous les termes de H Pour la règle (1), la confiance vaut **[28] :**

$$conf\ [B \Longrightarrow H] = \frac{nombre\ de\ documents\ v\acute{e}rifiant\{t_1, t_2, t_3, t_4, t_5\}}{nombre\ de\ documents\ v\acute{e}rifiant\{t_1, t_2\}} \in [0,\ 1]$$

En probabilité, la confiance mesure la probabilité conditionnelle de H sachant B comme il est illustré dans la formule suivante: **[28]**

$$conf\ [B \Longrightarrow H] = P(H \mid B) = \frac{sup\ [B \Longrightarrow H]}{nombre\ de\ documents\ v\acute{e}rifiant\{t_1, t_2\}} \in [0,\ 1]$$

Lorsque la confiance vaut 1, la règle est dite **totale**. Dans le cas contraire, la règle est dite **partielle** à x %. **[28]**

4.2.2. Composition des règles d'association

De façon générale, une règle d'association se compose de deux parties : une première partie appelée antécédent (si) qu'on trouve toujours dans les données et une deuxième partie appelée conséquent (alors) qu'on obtienne après une combinaison avec l'antécédent.

4.2.3. Opérations des règles d'association

On peut effectuer deux opérations avec les règles d'association :

- **Calcul du Support de la règle** : c'est le nombre des transactions des éléments.
- **Calcul de la Confiance de la règle** : c'est le taux de présence d'Y quand X est présent au sein de la transaction. **[29]**

4.2.4. Étapes d'extraction des règles d'association

Le processus qui résume les étapes d'extraction des règles d'association se déroule en trois étapes **[26].** Mais comme le processus est à la fois itératif et interactif, le processus nécessite une phase de prétraitement des données avant tout et donc ramène le processus en quatre étapes selon Nicolas Pasquier **[30]** qui les illustre dans le schéma ci-après :

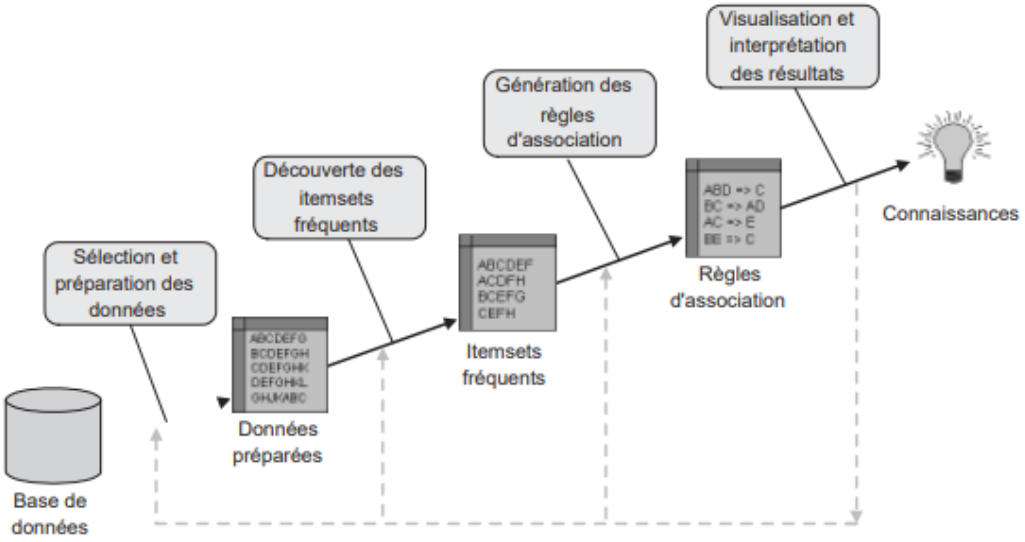

Figure 4.1. Étapes du processus d'extraction de règles d'association [37]

1. **Prétraitement**

C'est une étape préliminaire qui permet à l'utilisateur de connaitre les données et le domaine d'application afin de participer à la sélection et à la préparation des données. **[30]**

2. **Préparation des données**

On peut utiliser plusieurs types de bases de données (relationnelles, transactionnelles, spatiales, multimédia, temporelles, orientées objet) pour extraire les règles d'association. Pour cette première phase proprement dite du processus (excepter la phase préliminaire), on effectue deux étapes

principales, la sélection des données de la base qui permettront d'extraire les informations intéressant par l'utilisateur et la transformation de ces données en un contexte d'extraction. **[30]**

Cette étape consiste à réduire la quantité des données en gardant seulement celles qui sont plus pertinentes **[31], [26],** et en transformant par la suite ces derniers en un contexte d'extraction c'est-à-dire en les transformant en un ensemble d'objets, d'un ensemble d'ltemsets ainsi qu'une relation binaire entre les deux. Cette transformation des données en données binaires permettra d'améliorer la qualité des règles d'association. **[26]**

3. Recherche ou découverte des ItemSets fréquents

Étant donné un contexte d'extraction de règles d'association B, la découverte des itemsets fréquents est un problème non trivial, car le nombre d'itemsets fréquents potentiels est exponentiel en fonction du nombre d'items du contexte B. **[30]**

Un Itemset est dit fréquent, quand il est constitué d'un ensemble d'éléments dont le support est supérieur ou égal à un certain support minimal spécifié par l'utilisateur. Cette étape est très coûteuse en temps d'exécution. Pour un ensemble de n items par exemple, le nombre d'ltemsets fréquents qui peut être généré est de 2^n. **[26]**

4. Production des règles d'association

La génération des règles d'association consiste à déterminer les règles d'association dont le support et la confiance sont supérieurs ou égaux à un certains support et confiance minimaux définis par l'utilisateur. **[26]**

4.2.5. Critères d'évaluation des règles d'association

De manière générale, pour évaluer les règles d'association nous utilisons toujours les deux (2) critères de base : le support et la confiance. **[32]**
L'illustration des critères d'évaluation est expliquée dans l'exemple ci-après :
Soit la règle d'association R1 : Si p1 alors p2.

Données

Caddie	p1	p2	p3	p4
1	1	1	1	0
2	1	0	1	0
3	1	1	1	0
4	1	0	1	0
5	0	1	1	0
6	0	0	0	1

Tableau 4.1. Représentation des données binaires d'évaluation des règles d'association

Le support : C'est un indicateur de fiabilité de la règle qui fixe le seuil en dessous duquel les règles ne sont pas considérées comme fiables : **[32]**

- sup(R1) = 2 en termes absolus ou
- sup(R1) = 2/6 = 33% en termes relatifs

La confiance : elle représente un indicateur de précision de la règle qui permet de fixer un seuil de support minimum et de confiance. Plus la confiance est élevée, meilleure est la règle d'association. **[32], [33], [34].**

En utilisant l'algorithme Apriori, on se fixe souvent le seuil qui sera considéré comme support minimum et une confiance.

Puis suivront les étapes de calcul du support de chaque item où tous ceux qui seront inférieurs au seuil minimum seront supprimés de cette étape **[32].** Et après, on identifie tous les itemsets fréquents pour enfin calculer la confiance de chaque règle d'association qui en découle et nous ne conservons que celles qui satisfont aux critères de confiance. **[32]**

$$conf(R1) = \frac{\sup(R1)}{\sup(antécédent\ R1)} = \frac{\sup(p1 \rightarrow p2)}{\sup(p1)} = \frac{2}{4} = 50\%$$

NB. : Pour une bonne règle d'association, le support et la confiance doivent être élevés **[33].**

4.2.6. Exemple du Codage disjonctif complet

Nous allons nous servir d'un cas comme exemple pour illustrer les associations.

Observation	Taille	Corpulence
1	Petit	Mince
2	Grand	Enveloppé
3	Grand	Mince

Observation	Taille = petit	Taille = grand	Corpulence = mince	Corpulence = enveloppé
1	1	0	1	0
2	0	1	0	1
3	0	1	1	0

Figure 4.2. Exemple Codage disjonctif complet

Dès que l'on peut se ramener à des données 0/1, Il devient possible de construire des règles d'association. Il s'agit de détecter les cooccurrences des modalités (attribut = valeur). Certaines associations sont impossibles par construction (exemple : on ne peut pas être « petit » et « grand » en même temps) **[33]**

4.2.7. Algorithmes d'extraction des itemsets fréquents

4.2.7.1. Algorithmes Apriori et OCD

L'algorithme Apriori représente la base de tous les algorithmes de recherche des règles d'association. Il utilise une stratégie de recherche des Itemsets fréquents en commençant par les Items sets les plus généraux vers les plus spécifiques. **[26]**

Les algorithmes Apriori et OCD ont été créés de façon indépendante en 1994, mais procèdent de la même façon pour extraire les itemset, et les deux sont des algorithmes itératifs de recherche des itemsets fréquents par niveaux. Cela signifie que durant la k^{ieme} itération, un ensemble d'itemsets candidats de taille k est généré et un balayage du contexte est réalisé afin de supprimer les candidats infréquents. L'ensemble des k-itemsets fréquents ainsi générés est utilisé lors de l'itération k + 1 suivante pour générer les candidats de taille k + 1. Ces algorithmes réalisent donc u itérations afin de déterminer tous les itemsets fréquents dans l'ordre croissant de leur tailles, u étant la taille des plus grands itemsets fréquents. **[30]**

L'algorithme Apriori se base sur les deux propriétés suivantes :

1. Tous les sous-ensembles d'un itemset fréquent sont fréquents. Elle a été proposée pour permettre la limitation du nombre de candidats de taille k produite lors du $k^{ième}$ itération (parmi les 2k itemsets de taille k existants) en réalisant une jointure conditionnelle des itemsets fréquents de taille k-1 découverts lors de l'itération précédente. **[30]**
2. Tous les surensembles d'un itemset infréquenté sont infréquents. C'est pour supprimer un candidat de taille k lorsque au moins un de ses sous-ensembles de taille k-1 ne fait pas partie des itemsets fréquents découverts lors de l'itération précédente. **[30]**

La **Figure 4.2** illustre un algorithme utilisé dans l'extraction des règles d'association appelé algorithme Apriori

Input : un support minimum et une base de données de transactions

Output : génération des itemsets fréquents

1. $M_i = \Phi; i = 0$
2. $K_1 =$ tous les 1-itemsets dans la base de données
3. $F_1 =$ tous les itemsets fréquents de K_1

Tant que(M_{i+1} est non vide) Faire

1. $K_{i+1} =$ Candidate-gen(F_i)
2. $F_{i+1} =$ tous les itemsets fréquents de K_{i+1}
3. $i++$
4. Retourner l'union des M_i

Fin tant que

Figure 4.3. Présentation de l'algorithme Apriori [26]

4.2.7.2. Algorithmes Apriori Tid

Il est juste une variante de l'algorithme Apriori qui permet de réduire progressivement la taille du contexte afin de permettre un stockage en mémoire et ne plus réaliser d'opérations d'entrée/sortie, après le premier balayage de celui- ci.

4.2.8. Avantages et inconvénients des règles d'association

4.2.8.1. Avantages

1. Elles sont appliquées dans plusieurs domaines de la vie quotidienne, comme l'analyse du panier de la ménagère ;
2. Elles permettent de découvrir des connaissances utiles qui sont cachées dans des grandes bases des données ;
3. Elles sont simples à utiliser, efficaces et faciles de comprendre ;
4. Elles ont un formalisme non supervisé et général ;
5. Elles ont des résultats clairs et faciles à interpréter ; **[26], [31]**
6. Le forage des règles d'association est un grand succès dans divers domaines que ce soit dans des activités commerciales, sociales ou humaines. **[31]**

4.2.8.2. Inconvénients des règles d'association

1. On découvre dans le traitement des données plusieurs règles d'association dont la majorité n'est pas intéressante;
2. Le processus prend beaucoup de temps de recherche des Itemsets fréquents;
3. Les algorithmes utilisés ont trop de paramètres, ce qui rend complexe l'extraction des données surtout pour les non expérimentés en traitement des règles d'association;
4. Un problème de sécurité pourrait être posé: des renseignements confidentiels peuvent être facilement divulgués, en utilisant cette technique;
5. L'utilisation d'un seul minsup pourrait engendrer un dilemme d'item rare, celui-ci signifie que tous les éléments de la base de données sont de même nature. Ce qui n'est pas toujours vrai. **[31]**

4.2.9. Problématiques des règles d'association ordinaire

L'utilisation des règles d'association ordinaire ou régulière permet certes de dénicher beaucoup d'associations intéressantes, mais seulement elles ne sont pas capables de trouver les associations les moins fréquentes qui sont cachées à l'intérieur des corpus. **[35] [36]**

Avec les algorithmes qui sont utilisés dans l'extraction des règles d'association régulière, on remarque souvent une perte de certaines d'entre elles. On constate aussi que les mots qui sont strictement liés dans un ensemble de documents apparaissent fréquemment ensemble.

Exemple du mot imprimante qui peut apparaître souvent avec le mot papier. Dans ce cas, on aura des associations spécifiquement appropriées à un terme, mais pas à d'autres. Une association entre imprimante et encre aurait une confiance basse, puisqu' il y a beaucoup de transactions où le mot imprimante est lié au mot papier (sans le mot encre). **[35]**

C'est pour cette raison que les règles d'association ont été améliorées en créant les règles d'association maximale que nous allons expliquer dans le paragraphe suivant et qui seront utilisées dans notre solution.

4.3. Règles d'association maximale

Elles représentent un outil efficace pour dégager ~~des~~ d'intéressantes relations dans un corpus en enlevant les relations les moins intéressantes qui ne peuvent pas être capturées par des règles d'associations ordinaires.

Pour mieux étudier la qualité des règles et éliminer les règles d'association redondantes, on y ajoute des solutions statistiques comme la régression linéaire ou logistique.

4.3.1. Définition des concepts

4.3.1.1. Règles d'association maximale

Les règles d'association maximale ont fait leur apparition dans le début des années 60. C'est un type de règles d'association représentant une méthode très pertinente pour extraire des relations existantes dans un ensemble de données. Une règle d'association maximale est notée comme suit : X $MAX\Rightarrow$ Y **[37].** Ce sont des moyens très efficaces qui permettent de contourner les problèmes liés à la découverte des associations les moins fréquentes qui sont cachées à l'intérieur des corpus, car on est capable d'extraire des relations les moins intéressantes qui ne peuvent pas être capturées par des règles d'associations ordinaires. **[35]**

4.3.1.2. M-Support

On note le support maximal d'un item X par Smax (X). Soit Ti une catégorie de I, avec X un itemset tel que X ⊆Ti, on dit que X apparait seul dans une transaction t, si et seulement si Ti =X. autrement dit, X est le plus grand sous-ensemble de Ti qui est dans la transaction t.

On dit qu'une règle X ⟹ Y M-supporte un item X, si X apparaît seul dans une transaction t alors l'item Y apparaît également.

Le M-support représente le support maximal de la règle X $MAX\Rightarrow$ Y.

Le M-support d'une règle X $MAX\Rightarrow$ Y représente toutes les transactions de la base des transactions que M-Supporte X et Supporte Y. **[37]**

$$\text{M-Support } (X \overset{MAX}{\Longrightarrow} Y) = \{ t \in T \text{ tel que } t : M - \text{supporte X et supporte Y}\}$$

4.3.1.3. M-Confiance

Le M-confiance représente la confiance maximale de la règle X $MAX\Rightarrow$ Y. Elle est notée par Cmax(X $MAX\Rightarrow$ Y).

Soit T(Y) la catégorie d'Y, on note D(X, T(Y)) le sous-ensemble de la base de données D constitué de toutes les transactions qui M-supportent X et qui contiennent au moins un élément de T(Y). **[37]**

La formule de M-Confiance sera la suivante :

$$\text{M-Confiance } (X \overset{MAX}{\Longrightarrow} Y) = \frac{\text{M-Support } (X \overset{MAX}{\Longrightarrow} Y)}{D(X,T(Y))}$$

4.3.2. Propriétés

Les règles d'association maximale possèdent les mêmes propriétés que les règles d'association ordinaires (les règles ne peuvent être combinées, décomposées, elles ne sont pas transitives, les items de chaque transaction sont uniques, les items sont inclus dans l'ensemble de départ, les sous-ensembles sont disjoints). La différence résulte dans la manière de calculer le support de la règle ainsi que la confiance. **[29]**

4.3.3. Opérations

Le calcul des règles d'association maximales est en général beaucoup plus rapide que le calcul d'associations régulières (à condition que le nombre de catégories ne soit pas trop grand). La raison en est que pour chaque catégorie, toute transaction M-prend en charge au plus un ensemble d'éléments.

Ainsi: **[36]**

1. Tous les ensembles M-fréquents peuvent être générés en un petit nombre de passages sur la base de données ;
2. Pour chaque catégorie, les ensembles M-fréquents cloison la base de données en petites sous-bases de données. **[35]**

Calcul du support maximal de la règle : Le support maximal traduit le calcul du nombre de transactions satisfaisant à la définition de la règle d'association maximale ;

Calcul de la confiance maximale de la règle : c'est le pourcentage de transactions satisfaisantes à la définition de la règle d'association maximale par rapport au nombre de transactions contenant au moins un élément de l'ensemble **[29].**

4.3.4. Conclusion sur les règles d'association

D'une façon générale, les règles d'association maximum présentent une efficacité plus que les règles d'association régulière, car elles sont capables de ressortir les règles d'association dont la pertinence est plus ou moins importante, mais qui peuvent par la suite devenir importantes dans la compréhension d'un corpus.

Mais il est important de retenir que les règles d'association maximale n'ont pas été créées pour remplacer totalement celles dites régulières, mais elles ont été créées pour les compléter **[35]**

4.4. Régression logistique

La régression logistique est l'un des modèles d'analyse multivariée couramment utilisés en épidémiologie. C'est grâce à elle qu'on peut mesurer l'association entre la survenue d'un évènement (variable expliquée qualitative) et les facteurs susceptibles de l'influencer (variables explicatives). **[38]**

Elle est beaucoup utilisée dans la prédiction des variables binaires appelées aussi variable de Bernoulli qui ne peut être pour notre cas que 0 (fausse attaque) ou 1 (vraie attaque). Elle caractérise les relations entre une variable dépendante (ou variable à expliquer) et une seule (régression logistique simple) ou plusieurs variables prises en compte simultanément (régression logistique multiple). C'est donc un modèle permettant de relier la variable dépendante (Y) à des variables explicatives (X1 , X2 , X3 , … Xi) dans le but de mesurer l'association entre la survenue d'un évènement (variable expliquée qualitative) et les facteurs susceptibles de l'influencer (variables explicatives). **[38]**

4.4.1. Définition et solution possible

La régression logistique est une technique statistique qui permet d'ajuster une surface de régression a des données lorsque la variable dépendante est dichotomique. Elle est plus utilisée pour mener les études dont le but est de vérifier si les variables indépendantes peuvent prédire une véritable dépendante dichotomique. **[39]**

Il existe trois (3) solutions possibles permettront d'évaluer la régression logistique :

1. **Les courbes ROC**, ce sont des représentations graphiques de la relation existante entre la sensibilité et la spécificité d'un test pour toutes les valeurs seuils possibles. L'ordonnée du graphique représente la sensibilité et l'abscisse correspond à la quantité.**[40]**
2. **Évaluation du R^2 de McFadden** qui sert à évaluer la proportion de la variance expliquée à travers un modèle par rapport à la variance totale des données. Plus que R^2 est élevé, plus la proportion de la variance expliquée par le modèle est grande. Cette approche est la plus utilisée.
3. **Tests d'hypothèses** c'est une énonce qui permet de valider si les coefficients du modèle (bi) sont significatifs. **[41]**

4.4.2. Contexte

Comme nous l'avons explicité précédemment, la régression logistique une méthode statistique de prédiction des classes binais. Le résultat ou la variable cible qui est produit dans cette méthode est de nature dichotomique c'est-à-dire avec la possibilité de deux (2) classes. **[42]**

La régression logistique est possible dans les contextes suivants :

- Y qui est la variable explicative est une variable binaire (0 ou 1) ;
- 0 en cas de non-occurrence de l'évènement ;
- 1 si occurrence de l'événement ;
- Si Y est aléatoire et Xi est non aléatoires ;
- On cherche à expliquer la survenue d'un évènement ;
- On cherche la probabilité de succès ;
- On travaille en termes d'espérance. **[41]**

4.4.3. Utilisation de la régression logistique

La régression logistique est utilisée dans plusieurs cas de figure liés aux problèmes de classification comme la détection des intrusions dans un système.

La régression logistique est l'un des algorithmes d'apprentissage automatique les plus simples et les plus couramment utilisés pour la classification à deux classes. Non seulement que sa mise en œuvre est facile, elle peut être aussi utilisé comme référence pour tout problème de classification binaire. Ses concepts fondamentaux de base sont également constructifs en apprentissage profond. Elle décrit et estime la relation entre une variable binaire dépendante et des variables indépendantes. **[42]**

4.4.4. Types de régression logistique

Nous pouvons selon **Avinash Navlani** retenir trois (3) types de régression logistiques :

- **Régression logistique binaire ou ordinaire** : ici, la variable cible n'a que deux résultats possibles par exemple s'il y a attaque ou pas attaque.
- **Régression logistique multinomiale** : la variable cible, qualitative a trois catégories nominales ou plus.
- **Régression logistique ordinale** : la variable cible, qualitative comporte au moins trois catégories ordinales. **[42]**

4.4.5. Modèle de régression logistique

En s'appuyant sur le modèle de régression logistique simple et multivariée, nous allons dans notre travail utiliser celle dite multivariée, car nous disposons, dans notre dataset de plusieurs variables explicatives (X) prédictives.

Étant donné que nous sommes dans la régression logistique binaire où Y représente la valeur de la variable dépendante dichotomique qui ne peut prendre que la valeur 0 pour présenter le manque d'attaque, l'échec ou le « non », ou la valeur 1 pour présenter la présence, le succès ou bien le « oui », tandis que X représente les valeurs des différents attributs prédictifs relatifs à chaque échantillon ou participant pouvant avoir des valeurs discrètes ou continues. **[43]**

Pour se faire, l'équation de régression logistique est représentée comme suite : **[43]**

$$\mathbb{P}(Y=1|X)=\frac{e^{\beta 0+\beta 1X1+\beta 2X2+\beta 3X3+\cdots+\beta nXn}}{1+e^{\beta 0+\beta 1X1+\beta 2X2+\beta 3X3+\cdots+\beta nXn}}$$

Avec :

P(Y=1|X) = la probabilité de la modalité pour laquelle le modèle est construit

Y = variable à expliquer ou variable critère

X= variables explicatives

Normalement, la régression logistique donne toujours un résultat qui se situe entre 0 et 1 **[44]**

Nous aurions les cas ci-après :

- Quand il est égal à 0, la probabilité que l'événement arrive est nulle
- Quand il est proche de 0, la probabilité que l'événement arrive est faible
- Quand il est proche de 1, la probabilité que l'événement arrive est élevée
- Et quand c'est égal à 1, la probabilité que l'événement arrive est parfaite

4.4.6. Interprétation

Nous pouvons interpréter un modèle de régression logistique en termes de valeur de coefficients de régression ou en termes d'odds ration (rapport de cotes). **[45]**

Parlons de l'interprétation des résultats de la régression logistique, il faut retenir que si avec la régression linéaire il est juste question de minimiser la somme des carrées d'erreur, ici on aura une fonction dite de régression logistique.

On procède au choix des coefficients de la régression logistique qui se repose plutôt sur l'obtention des valeurs prédites de Y situées le plus près possible des valeurs observées. Ces coefficients constituent les paramètres d'estimation de la probabilité maximale (maximum-likelihood) et mesurent le changement du ratio de probabilité (odds ratio). **[44]**

4.4.6.1. Niveaux d'analyse du modèle

Nous pouvons évaluer la qualité globale d'un modèle et l'utilité de chacun de ses éléments dans les régressions linéaire et logistique. Cependant, les coefficients du modèle peuvent soit être interprétés directement soit être transformés en odds ratio (rapport de cote). Cette seconde forme est plus naturelle dans le cas de la régression logistique. **[45]**
[44]

4.4.6.2. Qualité globale du modèle : déviance

Une régression qu'elle soit linéaire ou logistique a comme objectif premier de faire à ce que la variable ajoutée au modèle permette plus d'efficacité de prédire l'appartenance au groupe. **[44]**

La déviance encore appelée -2 Log Likelihood ou -2LL représente la distance qui sépare le modèle et les observations. Son rôle est de séparer le modèle basique n'a aucune variable explicative et le modèle complet où il y a toutes les variables.

La distance entre le modèle et les observations est indiquée par le chiffre (2), et il sied de retenir que plus cette distance est faible, plus le modèle est meilleur ou bon. Si on n'a aucune variable explicative et si on prend toutes ces explications du modèle précédent des variables indépendantes, la déviance est plus petite et donc le second modèle est meilleur que le modèle où on a juste les intercepte. **[45]**

4.4.6.3. Qualité globale du modèle : Chi-2

La statistique chi-2 permet d'évaluer l'amélioration de la déviance par rapport au modèle « NULL » (avec constante seulement). En réalité, si la statistique du chi-2 n'est pas significative, le modèle est rejeté. Cette statistique permet aussi de comparer des modèles entre eux. Un modèle doit obligatoirement être accepté et cette acceptation ne veut pas dire que le modèle est bon. C'est aussi grâce à cet outil qu'il est possible de valider un modèle. **[45]**

4.4.6.4. Qualité globale du modèle : pseudo-R^2

Représentés parfois par Cox et Snell, les R^2 correspondent à la mesure de l'utilité des variables explicatives. Bien qu'ils soient des estimations, mais leur utilisation ne pas la même comme c'est le cas dans la régression linéaire.

4.4.6.5. Test individuel des variables

Le test individuel de chaque variable permet de déterminer la significativité des variables. Grâce à la significativité des coefficients que l'on saura si la variable correspondante est significative, si non, l'on prendra la décision de la supprimer du modèle.

4.4.6.6. Variables explicatives à plus de 2 modalités

Lorsqu'une variable explicative catégorielle comporte plus de 2 modalités, elle est remplacée dans le modèle par plusieurs variables muettes. Un p-valeur est fourni pour chacune de ces variables muettes. Toutes ces variables muettes correspondent à la même variable explicative. Il faut donc soit toutes les laisser dans le modèle soit toutes les supprimer. Il est possible de tester globalement l'effet de toutes les variables muettes correspondant à une même variable explicative en utilisant la statistique du chi-2 pour comparer le modèle avec et sans ces variables muettes. **[45]**

4.4.6.7. Avantages et désavantages

4.4.6.7.1. Avantages

La régression logistique possède plusieurs avantages comme :

- Sa nature qui reste très efficace et simple,
- Elle n'a pas besoin de puissance de calcul élevée,
- Elle est facile à mettre en œuvre,
- Son interprétation est facile,
- Elle est beaucoup utilisée par les analystes de données et les scientifiques.
- Elle ne nécessite pas de mise à l'échelle des fonctionnalités.
- La régression logistique fournit un score de probabilité pour les observations **[42]**

4.4.6.7.2. Désavantages

Malgré les nombreux avantages que présente la régression logistique, il a aussi des désavantages.

- Elle n'est pas capable de gérer un grand nombre de caractéristiques ou variables catégorielles ;
- Elle est vulnérable au surajustement ;
- Elle ne peut pas résoudre le problème non linéaire avec la régression logistique ;
- La régression logistique ne fonctionnera pas bien avec des variables indépendantes qui ne sont pas corrélées à la variable cible et qui sont très similaires ou corrélées les unes aux autres. **[42]**

4.4.6.8. Conclusion sur la régression logistique

La régression logistique est une méthode d'analyse multivariée puissante permettant d'obtenir une quantification de l'association entre une variable et chacun des facteurs l'influençant la variable, tout en tenant compte de l'effet simultané des autres facteurs. Elle permet ainsi de contrôler de possibles biais de confusion. Son emploi est rendu aisé par l'utilisation de logiciels statistiques. **[38**] C'est donc une solution très efficace dans la modélisation, car elle est capable de produire avec précision les variables les plus significatives.

Dans le chapitre suivant, nous présenterons quelques solutions proposées par certains auteurs sur l'amélioration les IDS et l'utilisation de la statistique, des règles d'association, de la régression logistique dans l'optimisation des IDS.

CHAPITRE 5 : ÉTAT DE L'ART

5.1. Introduction

La sécurité des véhicules électriques dans le réseau V2G demeure une préoccupation majeure des opérateurs et motive les chercheurs à trouver des solutions afin d'améliorer leur niveau de sécurité. Si dans les chapitres précédents, nous avons présenté des généralités sur le V2G, les systèmes de détection d'intrusion, les règles d'association et la régression logistique, dans celui-ci, nous allons faire une revue de littérature sur les recherches qui ont été faites concernant l'amélioration des systèmes de détection d'intrusion.

Plusieurs solutions qui utilisent des algorithmes différents ont été proposées pour tenter d'améliorer la sécurité des réseaux en général et des systèmes de détection d'intrusion en particulier. Parmi ses solutions, certains utilisent un seul algorithme tandis que d'autres en associent plusieurs dans le but de satisfaire les consommateurs des services du V2G.

5.2. Systèmes de détection d'intrusion

Nous allons présenter dans cette partie de notre mémoire, les recherches qui ont été faites sur les systèmes de détection d'intrusion afin de comprendre les méthodologies qui ont été utilisées, les faiblesses et les points forts de leurs solutions afin d'avoir des idées sur les performances de notre solution.

Dans **[46],** les auteurs considèrent les systèmes de détection d'intrusion comme un des éléments fondamentaux de la sécurité d'un réseau informatique et les placent sur la première ligne de défense. La solution qu'ils ont proposée est un modèle qui consiste à faire une analyse approfondie d'apprentissage considéré comme un réseau de neurones à plusieurs couches entièrement empilées et connectées les unes aux autres. Ce réseau de neurones permet la mise en œuvre d'un système de détection d'intrusion qui est basée sur les flux capables de faire la classification multi-classes.

Ils ont obtenu les résultats à la suite d'une expérimentation faite avec une base de données d'évaluation de détection d'intrusion de la CICIDS version 2017 qu'ils ont mis à jour. Ces résultats ont montré que le modèle proposé peut obtenir des résultats prometteurs sur la classification multi-classes en ce qui concerne la précision des taux de détection et au taux de faux positifs (faux taux d'alarme) sur cet ensemble de données spécifique. Cette solution qui utilise l'ensemble de données séquentielles peut présenter quelques difficultés dans la répartition des données d'entrée (paramètres des défauts)

qui n'est pas souvent uniforme, car l'uniformité est recherchée dans l'espace des sorties du réseau de neurones.

Les auteurs de **[47], Subba Basant, Biswas Santosh et Karmakar Sushanta** soutiennent dans leur article que les données d'entrée qui sont souvent à l'entrée des systèmes de détection d'intrusion sont redondantes, surchargent les calculs et peuvent émettre les fausses alertes. C'est pour cette raison qu'ils proposent un modèle capable de minimiser les surcharges de calcul en utilisant la méthode de l'Analyse en composantes principales (ACP) qui va réduire le volume des données d'entrée sans perdre aucune information significative contenue dans les données d'origine.

Ils ont utilisé le jeu de données NSL-KDD de l'Institut canadien pour la cybersécurité de l'Université du Nouveau-Brunswick **[48]** pour analyser la performance du système de détection d'intrusion proposée. Ce jeu de données se compose de 41 entités qui ont été réduites par la méthode statistique d'Analyse en composantes principales, mais en conservant une variance de 85 % à 98 % de la base de données originales. Bien que les résultats semblent être satisfaisants, mais la base de données utilisée présente des lacunes telles que le nombre trop élevé d'enregistrements redondants qui fait que les algorithmes d'apprentissage soient biaisés vers les enregistrements fréquents, et les empêchent ainsi d'apprendre des enregistrements peu fréquents qui sont généralement plus nocifs pour les réseaux. **[48]** Cette situation peut également biaiser les résultats de l'expérimentation.

5.3. Règles d'association

Pour lutter contre l'attaque des réseaux, certains auteurs ont utilisé les règles d'associations pour optimiser les systèmes de détection d'intrusion.

Selon les études faites par **Hyeok Kong, Cholyong Jong and Unhyok Ryang** dans **[34]**, ils proposent un nouvel algorithme d'exploration de règles d'association pratique pour la détection d'intrusions dans les systèmes de détection d'intrusion (IDS). Ils utilisent l'extraction des règles d'association rares parmi les ensembles d'éléments peu fréquents pour améliorer les systèmes de détection d'intrusion. Ils développent un algorithme pratique à partir d'une base de données de paquets réseau qui permet l'exploration de règles d'association rare basées sur le hachage pour trouver des modèles d'anomalies rares, mais utiles à l'utilisateur. Cet algorithme peut être appliqué aux champs nécessitant de miner des modèles cachés qui sont rares, mais précieux. Les auteurs se sont basés sur la méthode de hachage pour se servir des ensembles d'éléments peu fréquents, qui peuvent présenter des avantages évidents de problèmes de limitation de la vitesse et d'espace mémoire sur l'association traditionnelle algorithmes d'exploration de règles. **[34]** Leur solution ne peut fonctionner qu'avec les

items les moins fréquents afin de gagner en vitesse et dans la limitation de l'espace mémoire à la différence d'autres règles d'association qui utilisent les items les plus fréquents. Cette solution peut entraîner la perte des certaines informations utiles qui se trouvent dans les items fréquents.

Dans l'étude **[49],** les auteurs s'appesantissent sur les techniques de détection d'intrusion de base et se concentrent sur la façon d'appliquer les règles d'association à la détection d'intrusion. Ils ont mis en œuvre un à l'algorithme de minage des règles d'association afin d'explorer des règles d'association capable de détecter des anomalies dans les systèmes de détection d'intrusion. Leur solution optimise l'algorithme de minage des règles d'association, et utilise la logique floue pour améliorer les performances du système.

5.4. Régression logistique

La régression logistique constitue une des méthodologies statistiques efficaces dans la modélisation des systèmes. On l'utilise aussi en réseau pour optimiser les systèmes de détection d'intrusion.

Dans **[43]**, l'auteur, rappelle le rôle des IDS (Intrusion Detection System) et propose une solution de prévision mathématique pour optimiser la détection et la prévention d'attaques. Il s'appuie sur le modèle probabiliste innovateur basé sur la régression logistique. Grâce à cette méthode, on peut estimer l'occurrence d'un événement (une attaque) en fonction des connaissances acquises préalablement. Elle se base sur un historique d'une base de connaissance (BC) qui permet d'estimer les paramètres de la régression logistique. La base est ensuite implantée dans le RSU. Lorsque le modèle de régression est validé, il est utilisé pour estimer la probabilité d'une attaque et si cette dernière est supérieure au seuil fixé à l'avance (50%), l'attaque est corroborée. Cette solution qui est pour autant efficace, mais présente plusieurs failles, telles que les données de la base de connaissance qui sont insuffisantes et le manque de scénario commun pour évaluer les différentes alternatives du modèle.

5.5. Ensemble de plusieurs algorithmes pour optimiser les IDS

Pour certains chercheurs, l'utilisation d'un seul algorithme ne suffit pas pour améliorer les systèmes de détection d'intrusion. Il en faut plusieurs pour obtenir des résultats plus efficaces et précis.

Dans l'étude **[50], Subba Basant, Biswas Santosh et Karmakar Sushanta** proposent un modèle capable d'identifier une intrusion dans un réseau, de déclencher une alarme chaque fois qu'il y a un comportement anormal dans le réseau qui est en dehors des limites de l'ensemble d'entraînement est observé. Les auteurs utilisent deux méthodes statistiques différentes qui sont : l'analyse discriminante linéaire (LDA) et la régression logistique (LR) pour améliorer les systèmes de détection d'intrusion. Dans leur solution, des nouveaux modèles efficaces des systèmes de détection d'intrusion en termes de la capacité de calcul significatif rendent plus adapté le déploiement en temps réel de surveillance et d'intrusion dans le réseau. Ils ont utilisé la base de données de référence NSL-KDD pour faire l'expérimentation du modèle afin de comparer avec les autres modèles et ont obtenu des résultats qui montrent que les performances en termes de précision et taux de détection de l'ensemble des modèles LDA et LR sont équivalentes et, dans certains cas, même meilleures que les autres modèles des systèmes de détection d'intrusion. Ces modèles fonctionnent très bien sur les problèmes de classification des classes binaires et multiclasses. Cependant, ce jeu des données utilisé présente souvent un nombre important de redondances qui pet biaiser les résultats.

En cherchant toujours à combiner plusieurs algorithmes pour améliorer un modèle, **Ozge Yucel Kasap, Nevzat Ekmekci et Utku Gorkem Ketenci** proposent dans **[51]** une solution qui permet aux clients de faire un bon choix des produits en utilisant un outil d'exploitation des données appelé PROPCA (Proximus Optimum Canistro). Les auteurs de l'article combinent conjointement deux (2) méthodes dans leur solution **:** l'analyse de la régression logistique et l'exploration de règles d'association pour faire des recommandations en marketing. Leur solution a selon eux, donné des meilleurs résultats par rapport aux mêmes algorithmes quand ils sont utilisés seuls. Pendant que les règles d'association se chargent de chercher les règles avec les données pertinentes, la régression logistique pour sa part se charge de prédire la probabilité d'un produit par un client. La combinaison de ces deux approches a été testée sur un ensemble de données bancaires réelles et a donné des bons résultats. Cette solution qui donne des bons résultats n'a pas été expérimentée avec une base de données des systèmes de détection d'intrusion.

Dans [52], Pullagura et Anuradha proposent un modèle ensemble pour le système de détection d'intrusion. Pour elles, les services internet sont à la base des vulnérabilités constatées dans les systèmes informatiques et rendent sa sécurité difficile. Elles proposent une solution avec une architecture efficace du système de détection d'intrusion qui utilise un modèle de classification facile capable de générer des taux bas de fausses alertes. Elles fondent leur modèle sur un ensemble basé sur deux algorithmes : un ensemble flou de sélection de fonctions (FEFS) qui est une unification de cinq (5) scores de fonction et la fusion de plusieurs classificateurs (FMC) qui est une agrégation effectuée par union floue d'opérations à trois classificateurs. Le système proposé par les auteurs a été

expérimenté sur un jeu de données KDD99 et présente un fonctionnement supérieur aux anciennes méthodes comme Vector Machines (SVM), K-Nearest Neighbor (KNN) et les réseaux de neurones artificiels (ANN). Il atteint 0,9, 0,95, 0,96 et 0,9 de précision. Le modèle proposé n'inclut pas des méthodes d'apprentissage automatique capable de détecter les nouvelles attaques ou invisibles.

5.6. Conclusion

Dans ce chapitre, nous avons présenté quelques recherches faites au sujet de l'amélioration des systèmes de détection d'intrusion avec l'utilisation des diverses méthodologies. Les solutions proposées par certains et même expérimentées par d'autres ont été faites avec des jeux des données de référence qui peuvent biaiser les résultats.

Si certains auteurs ont utilisé un seul algorithme, d'autres ont préféré mettre ensemble deux (2) algorithmes afin de rendre la solution plus efficace.

.

Comme il y a des IDS qui génèrent des alertes qui peuvent être des vrais positifs ou des vrais négatifs, nous allons proposer un modèle dans le chapitre suivant qui sera capable de réduire l'impact des fausses alertes dans le réseau V2G, de prendre en charge les alertes du réseau et de prendre des décisions pour gérer les informations de sécurité dans les réseaux V2G.

CHAPITRE 6 : MODÈLE PROPOSÉ

6.1. Introduction

Le contenu de ce mémoire étant basé sur la recherche de solution permettant d'améliorer les systèmes de détection d'intrusion dans les réseaux V2G (Vehicle-To-Grid), c'est la raison qui justifie la présentation descriptive des domaines en rapport avec notre contenu, dont le V2G, les IDS (Intrusion Detection System), les règles d'association maximale et la régression logistique.

Comme le réseau véhiculaire V2G fait souvent l'objet de beaucoup d'attaques qui sont capables de le perturber et bloquer la liaison entre le VE (Véhicules Électriques) et la borne de recharge, nous avons pensé apporter une solution qui améliorerait le système de détection des intrusions en utilisant les règles d'association maximale qui seront appuyées par d'autres outils (statistiques et informatiques) comme : le logiciel SPSS (Statistical package for the social sciences), le test de corrélation de Pearson et la régression logistique qui fera la modélisation.

Dans ce chapitre, nous présentons un nouveau modèle de détection des intrusions que nous allons expérimenter en utilisant une base de données que nous appellerons **Dataset.** La base de données sera générée via des simulations puis traitée avec des méthodes statistiques afin de produire une nouvelle approche avec une solution capable de bloquer les vraies attaques en générant automatiquement des alarmes et d'archiver les fausses attaques avec une probabilité presque nulle d'émettre des fausses alertes.

6.2. Points sécuritaires

Les points sécuritaires sont définis selon des critères de choix qui s'accommodent au système de sécurité que nous voulons mettre en œuvre. Les critères de choix sur lesquels nous allons nous baser sont :

- **Fiabilité** : c'est la capacité qu'a un IDS de bloquer toutes les attaques que l'infrastructure subit et de justifier les alertes qu'elle génère.
- **Réactivité** : c'est la capacité d'un IDS à détecter les nouvelles attaques de la manière la plus rapide possible et de faire régulièrement les mises à jour.
- **Facilité de mise en œuvre et l'adaptabilité** : c'est la facilité de mise en place d'un IDS et sa capacité à s'adapter au contexte où il doit fonctionner.
- **Performance** : c'est la capacité d'un IDS nouvellement déployée à ne pas affecter la performance du système à surveiller ou à sécuriser.

- **Multicanal** : c'est la capacité d'un IDS à exploiter plusieurs canaux pour émettre les alertes (email, pager, téléphone, fax...) pour plus de garantie et de donner le maximum d'informations sur l'attaque détectée en vue de préparer une réaction qui correspond au type d'attaque. **[53]**

Les points sécuritaires que nous allons mettre en œuvre vont répondre non seulement aux critères de choix, mais aussi aux objectifs de la sécurité telle que :

- **La disponibilité** : le système de détection d'intrusion que nous allons mettre en place ne doit pas perturber ou empêcher le processus de la connexion ou de la recharge des VE. Il doit servir seulement de contrôler, surveiller les nœuds qui vont se connecter à la borne.
- **L'authentification** : Notre IDS doit nous assurer que seuls les VE qui sont reconnus comme faisant partie du réseau véhiculaire V2G ont le droit de se connecter à la borne.
- **La confidentialité** : elle permettra à notre IDS de protéger le système contre les accès non autorisés.
- La **non-répudiation** : elle va permettre aux différents clients (VE) de reconnaitre toutes les transactions pour lesquelles ils seront responsables. Aucun d'eux ne pourra nier les activités qu'il posera dans le réseau. **[53]**

6.3. Méthodologie

Dans le contexte de notre travail, nous aurions plus besoin des méthodes liées à l'expérimentation, la conception et la production en passant par une analyse. Elles vont déboucher vers la production de solutions très utiles et normalement très sophistiquées **[54].**

Suite à notre analyse du document de référence qui est la norme ISO 15118 et particulièrement, ses deux premières parties et étant donné que cette norme présente des limites, nous allons proposer un nouveau modèle à partir des règles d'association maximale et de la régression logistique. Le modèle sera expérimenté afin de produire une solution efficace et capable d'améliorer significativement les systèmes de détection d'intrusions du réseau V2G.

L'analyse portera aussi sur un dataset que nous avons simulé et que nous allons traiter afin de ressortir les variables indépendantes qui sont fortement corrélées avec la variable TYPE d'attaque qui montre s'il y a eu attaque de type déni de service ou homme du milieu. Après l'exploitation de la corrélation, la dimension de notre dataset sera réduite, car nous allons nous servir seulement des variables qui sont en corrélation avec le type d'attaque positif.

Une fois notre dataset réduit, il sera soumis au principe des règles d'association maximal afin de ne se servir que des variables ou items qui forment des associations les plus importantes avec la variable ATT (attaque) qui confirme s'il y a eu effectivement attaque ou pas. Notre dataset connaitra de nouveau une réduction dont la régression logistique se servira pour modéliser et obtenir des résultats du passage vers une expérimentation de tri des vraies attaques qui seront exploitées d'une part, et des fausses attaques d'autre part qui seront archivées.

6.4. Simulation de la base de données

6.4.1. Présentation de la base de données

Nous avons utilisé une base de données V2G préparée et générée à partir d'un simulateur des données appelé miniV2G.
Cette base de données au format xlsx a été obtenue sur base de trois (3) scénarios en utilisant l'outil Wireshark pour sniffer les interfaces réseaux. Ces scénarios sont :

- Scénario sans attaque ;
- Scénario avec attaque de type homme du milieu ;
- Scénario avec attaque de type DOS (Denial Of service).

Les données obtenues avec l'outil Wireshark ont été enregistrées sous le format "PCAP" alors que le CICFlowmeter qui est plus reconnu pour sa capacité à générer des datasets relatifs aux attaques a été utilisé pour générer des bases de données partielles en format Excel (les fichiers "PCAP" enregistrés ont été importés). On obtient ainsi, trois (3) fichiers Excel (base de données partielle) correspondant à chaque scénario.

Pour chaque base de données partielle les variables ID du flux, IP Source, IP destination, Port Source, Port Destination et les variables dont les données sont identiques pour les trois (3) fichiers Excel ont été supprimées. Et les trois (3) bases de données partielles nettoyées ont été fusionnées pour enfin obtenir une base de données finale qui comporte 25 variables comme le montre le **Tableau 6.1** qui décrit ces variables.

6.4.1.1. Construction du scénario

Dans un système de sécurité basé sur les IDS, la définition préétablie du déroulement de la procédure à exploiter se fait sous le nom de scénario. Cette phase est très capitale dans un modèle donné dans la mesure où il indique le plan global du système de sécurité (IDS pour notre cas) à mettre en place et permettre à l'administrateur du réseau V2G à prendre une décision adaptée au type de problème.

6.4.1.2. But du scénario

Le scénario de notre projet relate les activités ou étapes à réaliser dans le but :

- De planifier les activités de la procédure d'optimisation des IDS dans le réseau véhiculaire V2G,
- D'identifier les objectifs du projet,
- D'améliorer les solutions existantes,
- De formaliser les interactions entre les activités,
- Et de gérer l'ensemble de la solution de notre projet.

6.4.1.3. Préalable

Avant de définir le scénario de notre modèle, nous devrions d'abord connaitre :

- Qui attaque ?
- Quel est l'impact de cette attaque dans le réseau ?
- De quels types d'attaques s'agit-il ?

1. **Qui attaque ? :** La ressource qui attaque c'est est une hacker, pirate … que les auteurs dans **[55]** considèrent comme une ressource qui commet tout type d'infraction, de la contrefaçon à l'intrusion dans les systèmes informatiques. Les attaquants visent à obtenir l'accès à des privilèges aux ressources du réseau sans autorisation et même sans appartenir à cette organisation dans le but de nuire, de récupérer des informations sensibles des clients et de l'entreprise, mais parfois elles visent le fonctionnement même de l'entreprise ou du réseau.
2. **Quel est l'impact de cette attaque dans le réseau ? :** une attaque perpétrée dans un réseau a plusieurs impacts comme les pertes financières, la dégradation de la réputation de l'opérateur ou de l'entreprise, la destruction du système, perte de confiance des clients.
3. **De quels types d'attaques s'agit-il ? :** Il est très important de connaitre le type d'attaque qu'on est victime avant de prétendre bien se protéger. Raison pour laquelle, il faut connaitre l'ensemble des attaques que le réseau V2G (pour notre cas) est susceptible de subir.

6.4.1.4. Scénario

Notre scénario se résume autour du diagramme des types d'attaques ci-après :

- L'homme du milieu
- Le déni de service (DOS et DDOS)

6.4.1.4.1. Homme du milieu

Nous avons utilisé un algorithme approprié au réseau V2G pour simuler une attaque de type homme du milieu. Cet algorithme est celui qui a été aussi utilisé par SAIDOU DIOP **[53]**

```
Début
  Répéter pour chaque véhicule
        Définir T;
      Nouvelle Session();
      Tant que (Session.date + T) > Système.date
            Engager communication;
      fin tant que;
      Clore Session;
   Jusqu'à fin simulation;
Fin
```

6.4.1.4.2. Déni de service (DOS et DDOS)

Nous avons utilisé un algorithme approprié au réseau V2G pour simuler une attaque de type déni de service. Cet algorithme est celui qui a été aussi utilisé par SAIDOU DIOP **[53]**

```
Début
    Répéter pour chaque borne
        Définir R, T, i, Start;
        Start = Système.date;
        i = 0
       Tant que Système.date < ( Start + T )
            Nouvelle Session();
            i = i + 1;
           Tant que i <> R
                Engager communication;
            Fin tant que;
            Clore Session;
          Fin tant que;
      Jusqu'à fin simulation;
Fin
```

- R est un seuil de requêtes au-dessus duquel l'EVSE ne va plus accepter les demandes venant des utilisateurs non précédemment authentifiés,
- T est la période pendant laquelle EVSE n'accepte plus des requêtes.

6.4.2. Description des variables

Le tableau ci-dessous présente les 25 variables de notre jeu de données V2G que nous avons simulé. Toutes ses variables ont des données de type quantitative.

No	Attribut	Description	Type
1	Flow Duration	Durée d'écoulement	Quantitative
2	Tot Fwd Pkts	Total des paquets dans le sens direct	Quantitative
3	Flow Pkts/s	Débit des paquets qui est le nombre de paquets transférés par seconde	Quantitative
4	Flow IAT Mean	Temps moyen entre deux flux	Quantitative
5	Flow IAT Std	Temps d'écart type deux flux	Quantitative
6	Flow IAT Max	Temps maximum entre deux flux	Quantitative
7	Flow IAT Min	Temps minimum entre deux flux	Quantitative
8	Fwd IAT Tot	Temps total entre deux paquets envoyés dans le sens direct	Quantitative
9	Fwd IAT Mean	Temps moyen entre deux paquets envoyés dans le sens direct	Quantitative
10	Fwd IAT Std	Temps d'écart type entre deux paquets envoyés dans le sens direct	Quantitative
11	Fwd IAT Max	Temps maximum entre deux paquets envoyés dans le sens direct	Quantitative
12	Fwd IAT Min	Temps minimum entre deux paquets envoyés dans le sens direct	Quantitative
13	Fwd Pkts/s	Taille de l'écart type du paquet dans le sens direct	Quantitative
14	Bwd Pkts/s	Taille de l'écart type du paquet envoyé dans le sens opposé	Quantitative
15	Subflow Fwd Pkts	Le nombre moyen de paquets dans un sous-flux dans le sens direct	Quantitative
16	Active Mean	Temps moyen pendant lequel un flux était actif avant de devenir inactif	Quantitative
17	Active Std	Écart type temps pendant lequel un flux était actif avant de devenir inactif	Quantitative
18	Active Max	Durée maximale pendant laquelle un flux était actif avant de devenir inactif	Quantitative
19	Active Min	Temps minimum pendant lequel un flux était actif avant de devenir inactif	Quantitative
20	Idle Mean	Temps moyen pendant lequel un flux était inactif avant de devenir actif	Quantitative

21	Idle Std	Écart type temps pendant lequel un flux était inactif avant de devenir actif	Quantitative
22	Idle Max	Temps maximum pendant lequel un flux était inactif avant de devenir actif	Quantitative
23	Idle Min	Temps minimum pendant lequel un flux était inactif avant de devenir actif	Quantitative
24	ATT	Attaque	Quantitative
25	TYPE	Type d'attaque	Qualitative

Tableau 6.1. Description des variables

6.4.3. Présentation des outils utilisés

6.4.3.1. MiniV2G

MiniV2G c'est un logiciel qui fournit un environnement de test sur les différents scénarios V2G à ses chercheurs. **[2]** Dans le cadre de notre travail, nous l'avons utilisé pour simuler le chargement V2G. Il est développé sur base de Mininet pour prendre en charge le développement du réseau tout en implémentant RiseV2G pour simuler le processus de charge réel basé sur la norme ISO 15118. MiniV2G fonctionne dans un scénario émulé léger, construit sur le dessus de Mininet. L'approche de virtualisation permet de partager facilement la configuration du réseau et le code pour répliquer les expériences. **[2]** Il intègre aussi le RiseV2G qui permet de simuler le processus de la charge dans le réseau V2G.

6.4.3.2. Mininet

Mininet est un logiciel du type open source, un émulateur capable de créer un réseau virtuel réaliste, exécutant un noyau, un commutateur et un code d'application réels, sur une seule machine (VM, cloud ou native) en utilisant la commande **sudo mn**. **[56], [2]** C'est donc grâce à lui et à Wireshark que nous avons implémenté notre réseau virtuel V2G en produisant avec plus de précision les communications personnalisées de notre réseau.

Il est caractérisé par sa flexibilité, son applicabilité, son interactivité, son évolutivité, son réaliste et sa possibilité de se partager. **[57]**

Il comprend également une interface graphique accessible par la commande Miniedit permettant de créer, concevoir, configurer et tester une topologie de réseau par glisser-déposer. De plus, il est possible d'interagir avec chaque machine virtuelle à l'aide de l'interface de ligne de

commande fournie. **[2]** Cette interface peut servir aussi à insérer dans le réseau virtuel un EV et un EVSE.

Le Mininet est un logiciel qui est capable de créer et personnaliser des éléments SDN, de les partager avec d'autres réseaux et effectuer des interactions. Dans l'ensemble de ses éléments se trouvent les hôtes, les commutateurs, les contrôleurs et les liens. Le Mininet permet aussi de prototyper rapidement de grands réseaux sur un seul ordinateur. Il crée des réseaux définis par logiciel évolutifs à l'aide de mécanismes de virtualisation légers, tels que des processus et des espaces de noms réseau. Ces fonctionnalités permettent au Mininet de créer, d'interagir avec, de personnaliser et de partager rapidement les prototypes. **[58]**

6.4.3.3. RiseV2G

Le RiseV2G c'est un simulateur de la plateforme V2G qui permet d'établir et maintenir une communication entre le véhicule électrique et la borne de recharge tout en respectant les recommandations de la norme ISO 15118. Dans ce logiciel Open source sont incluses toutes les fonctionnalités liées à la recharge d'un VE et à la sécurité telle que définie par ISO 15118-2, des dossiers avec le code source des simulations. Le logiciel a été testé lors du symposium international semestriel sur les tests ISO 15118 et le système de charge combiné (CCS) et a abouti à une solution stable et fiable pour créer des produits conformes à la norme ISO 15118. **[2]**

Java est le logiciel qui a été utilisé pour développer le simulateur RISEV2G qui est aussi efficace dans la création des certificats pour le PKI (l'infrastructure à clé publique).

6.4.3.4. Wireshark

Le Wireshark qui était appelé Ethereal à sa création est une application qui est plus utilisée dans la capture et l'affichage des données qui sont transmises sur le réseau avec la capacité d'effectuer plusieurs opérations comme l'analyse par des codes sources et la lecture sur les paquets d'un réseau pour enfin les filtrer pour répondre aux besoins des utilisateurs. Ces paquets sont facilement visibles en temps réel, on peut aussi les analyser hors ligne. **[59]**

Il dispose d'une interface d'utilisation qui est très conviviale et qui permet d'afficher des données produites par les différents protocoles qui se trouvent dans les différents types des réseaux interconnectés. Wireshark prend en charge des dizaines de formats de fichiers de capture / trace, notamment CAP et ERF, que nous avons exploité dans notre travail.

Cette application est importante dans la résolution des problèmes des réseaux, dans le développement et le test des logiciels avec les fonctionnalités comme :

- compatibilité avec les systèmes d'exploitation UNIX et Windows,
- capturer des données de paquets en direct à partir d'une interface réseau,
- ouvrir les fichiers contenant des données de paquets capturées avec tcpdump/WinDump, Wireshark et de nombreux autres programmes de capture de paquets,
- importer des paquets à partir de fichiers texte contenant des vidages hexadécimaux de données de paquets,
- afficher les paquets avec des informations de protocole très détaillées,
- enregistrer les données de paquets capturées,
- exporter certains ou tous les paquets dans un certain nombre de formats de fichiers de capture,
- filtrer les paquets sur de nombreux critères,
- rechercher des paquets sur de nombreux critères,
- coloriser l'affichage des paquets en fonction des filtres,
- créer diverses statistiques. **[59]**

Les outils de décryptage intégrés vous permettent de visualiser les paquets cryptés pour plusieurs protocoles populaires, notamment WEP et WPA / WPA2. **[59]**

6.4.3.5. CICFlowmeter

CICFlowMeter est un générateur de flux de trafic réseau distribué par CIC pour générer 84 fonctionnalités de trafic réseau. Il lit le fichier PCAP, génère un rapport graphique des caractéristiques extraites et fournit également un fichier CSV du rapport. Il s'agit d'une application open source écrit en Java et téléchargeable depuis Github. Ses codes sources peuvent être intégrés à un projet, car il offre plus de flexibilité en termes de choix des fonctionnalités qu'on souhaite : calculer, ajout de nouvelles, et également un meilleur contrôle de la durée du timeout du flux. **[60] [61]**

6.5. Modélisation

La modélisation dans un système comme le nôtre (du V2G) consiste à concevoir une structure ou une architecture globale du système, définir le dynamisme de ses composants et l'organisation des données nécessaires à ce système.

Dans le cadre de notre travail, nous allons utiliser le modèle de la régression logistique (LR) qui nous permettra de trouver les facteurs liés au succès de détection des vraies attaques et prédire le taux de réussite de cette détection **[62]** afin de ressortir les variables significatives et un important taux de prédiction à partir d'un échantillon qui se trouve dans le dataset que nous avons simulé.

6.5.1. Modèle de prévision (techniques de modélisation) et algorithme d'apprentissage d'optimisation des IDS

Dans [4], les auteurs présentent les modèles ci-après qui permettent de modéliser les IDS :

- Le modèle par arbres de décision (data mining) pour le système de détection par anomalie ;
- Le modèle par les réseaux de neurones pour les systèmes de détection par anomalie ;
- Les modèles statistiques ou des réseaux de Petri pour le système de détection par anomalie. **[63]**

Dans **[13]**, **Philippe Biondi** lui va plus loin en répertoriant les modèles suivants :

- **Data mining** : dont l'utilisation repose sur l'extraction des modèles descriptifs des énormes volumes de données d'audit.
- **Agents mobiles** : ils sont capables de détecter les attaques, les unir, et découvrir les stratégies d'attaques distribuées. Leur architecture est très résistante aux attaques, ils ont un système de partage de connaissances pour divulguer les informations de l'attaque afin d'améliorer sa technique. Ils sont imprévisibles et ont une diversité génétique, car ils se comportent comme des agents autonomes.
- **Réseaux de neurones :** ils sont très rapides dans le traitement et sont aussi résistants aux informations incomplètes ou déformées. Ils sont capables de filtrer et sélectionner les parties suspectes dans les données d'audit.
- **Immunologie** : elle utilise un algorithme de sélection négative pour détecter les intrusions. Elle est conçue pour détecter et éliminer les intrus dans le système. Elle est souple, donc capable d'apprendre et de reconnaître de nouvelles infections.
- **Algorithmes génétiques** : ce sont des algorithmes qui peuvent évoluer facilement et être utilisés pour obtenir une solution approchée à un problème d'optimisation, lorsqu'il n'existe pas de méthode précise pour laquelle on ne connait pas encore de solution. On les utilise pour trouver de solutions en un temps raisonnable. Ils utilisent la notion de sélection naturelle et l'appliquent à une population de solutions potentielles au problème donné. **[13]**

Étant donné que le domaine informatique évolue et des nouvelles solutions sont proposées, nous allons utiliser les règles d'association maximale qui constitue également un modèle très efficace

comme nous l'avons souligné dans le chapitre 4 et la régression logistique qui fait partie des solutions statistiques très bien adaptées dans la modélisation.

6.5.2. But de la modélisation

Le système de détection d'intrusions dans le réseau véhiculaire V2G doit être modélisé afin de trouver une solution optimale contre les attaques dont il est victime. C'est grâce au modèle qui sera développé dans la modélisation que nous aurions le schéma architectural de notre système qui part d'un problème donné jusqu'à la solution comme le montre le schéma ci-après :

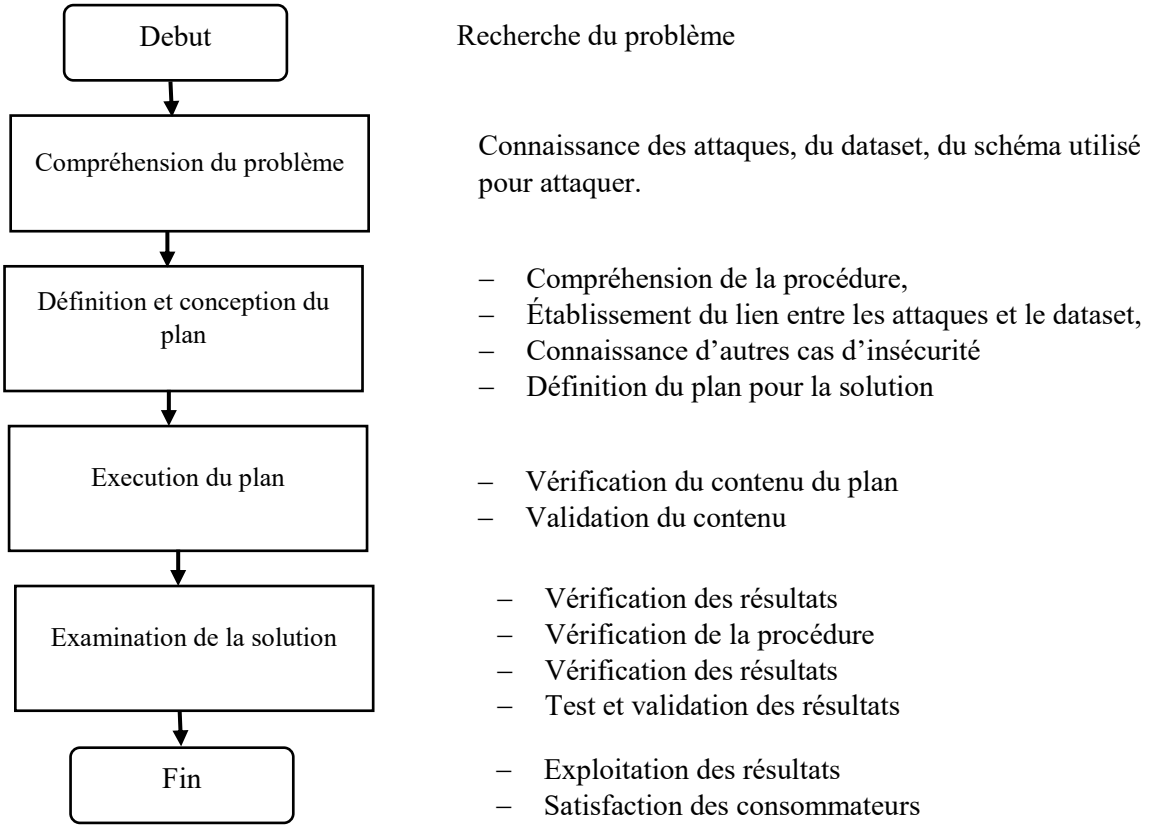

Figure 6.1. Les étapes du processus de la modélisation

Généralement, le modèle qui sort d'une modélisation doit converger vers une solution non seulement plus optimale que celles qui existent déjà, mais aussi plus fiable, rapide et indépendant.

6.5.3. Spécification des besoins

Pour améliorer le système de détection d'intrusion dans le réseau V2G, nous devrons passer par cette phase de la spécification des besoins qui nous permettra de mieux comprendre le contexte du système que nous voulons mettre en place. Nos besoins concernent des éléments ci-après :

- **Les acteurs du système** (opérateurs des bornes de recharge, utilisateurs ou VE, attaquants …)
- **Le scénario** qui retrace le contexte de notre travail, la procédure à suivre pour arriver à la solution …
- Et **les outils** qui sont constitués des méthodes statistiques (corrélation de Pearson, régression logistique, règles d'association) et des logiciels (langage de programmation Java, logiciel SPSS, logiciel YamBob et R)

6.5.4. Analyse des résultats

Cette analyse constitue une étape cruciale du traitement des données qui consiste à les préparer dans le but de nettoyer et transformer les opérations qui doivent être appliquées aux données brutes avant leur traitement, analyse et simulation. C'est grâce à cette étape que les données sont reformées et corrigées pour aboutir à un meilleur résultat.

6.5.4.1. Analyse de la corrélation de Pearson

Encore appelée corrélation linéaire simple, ou de Bravais-Pearson (ou de Pearson tout simplement), elle permet de caractériser une relation linéaire positive ou négative **[64].** Elle sera utilisée pour diminuer la dimension du nombre de variables de notre dataset dans le but de réduire la quantité des données à traiter avec la possibilité de garder les variables les plus utiles ayant un impact proche au type d'attaque et de supprimer les variables non pertinentes au système.

Le dataset réduit après l'analyse de la corrélation sera ensuite soumis à une analyse des règles d'association maximale pour enfin être modélisé par la régression logistique.

Nous tenons à signaler ici que pour l'ensemble des observations de départ, aucune d'elle n'a été supprimée pour des raisons de manque des données. Et donc notre corrélation va tenir compte des données traitées en considérant la corrélation significative à 0,01 et 0,05.

Nous avons exploité seulement la corrélation selon **Pearson** que nous allons interpréter sa matrice de corrélation obtenue à partir du logiciel statistique SPSS.

Pour confirmer la présence des corrélations entre les variables, trois (3) cas sont retenus :

- ($r > 0$ avec un maximum de +1) : on parle de la corrélation positive parfaite entre deux variables, plus le nombre augmente, plus le rendement augmente;

- (r < 0 avec un minimum de -1) : on parle de la corrélation négative parfaite, plus le nombre augmente donc on tend vers 0, plus le rendement diminue;
- (r = 0) : il y a absence de la corrélation, donc il n'y a aucune influence sur le rendement, aucune relation entre les variables.

Les résultats obtenus viennent de la corrélation bivariée entre les différentes variables de notre Dataset, et nous allons nous intéresser seulement à la corrélation entre la variable « Type d'attaque » et les autres dans le but de connaitre celles qui sont susceptibles d'augmenter la probabilité justifiante la présence d'une attaque dans le réseau V2G.

Nous avons obtenu dans l'ensemble un total de 12 variables qui sont en corrélation avec la variable Type d'Attaque et toutes les autres qui ne disposent d'aucune corrélation seront élaguées de notre dataset qui va finalement se retrouver avec 14 variables, car nous allons ajouter le **Type d'Attaque** et **Attaque**. Ces variables sont reprises dans le tableau ci-après :

No	**Attribut**	**Description**	**Type**
1	Flow IAT Mean	Temps moyen entre deux flux	Quantitative
2	Flow IAT Std	Temps d'écart type deux flux	Quantitative
3	Flow IAT Min	Temps minimum entre deux flux	Quantitative
4	Fwd IAT Tot	Temps total entre deux paquets envoyés dans le sens direct	Quantitative
5	Fwd IAT Std	Temps d'écart type entre deux paquets envoyés dans le sens direct	Quantitative
6	Fwd IAT Max	Temps maximum entre deux paquets envoyés dans le sens direct	Quantitative
7	Fwd IAT Min	Temps minimum entre deux paquets envoyés dans le sens direct	Quantitative
8	Fwd Pkts/s	Taille de l'écart type du paquet dans le sens direct	Quantitative
9	Bwd Pkts/s	Taille de l'écart type du paquet dans le sens opposé	Quantitative
10	Idle Mean	Temps moyen pendant lequel un flux était inactif avant de devenir actif	Quantitative
11	Idle Max	Temps maximum pendant lequel un flux était inactif avant de devenir actif	Quantitative
12	Idle Min	Temps minimum pendant lequel un flux était inactif avant de devenir actif	Quantitative
13	ATT	Attaque	Quantitative
14	TYPE	Type d'attaque	Qualitative

Tableau 6.2. Description des variables restantes

6.5.4.2. Analyse basée sur les règles d'association maximale

6.5.4.2.1. Présentation du logiciel

Comme indiquer ci-haut, le logiciel YamBob développé par **ABDOULAYE OUEDRAOGO [4]** est celui que nous avons utilisé pour extraire les règles d'association les plus importantes et élaguer celles qui ne nous serviront pas. Ce logiciel fonctionne avec une architecture basée sur six principaux modules ci-après:

- **Le module de récupération de données** qui consiste à importer les données de notre dataset (qui peuvent être au format TEXT, CVS, XLS ou XLSX). Le nôtre est au format XLSX compatible avec toutes les versions d'Excel.
- **Le module de traitement et de nettoyage** des données qu'on peut aussi appeler prétraitement consiste à traiter les données contenues dans le dataset et le nettoyer en élaguant les données manquantes, incorrectes ou incomplètes, car elles sont considérées dans notre cas comme des données non pertinentes. Au sortir de ce module, nous aurions un fichier Excel nettoyé avec toutes les valeurs correctes et complètes.
- **Le module de génération des itemsets fréquents** qui consiste à générer l'ensemble des règles d'association possible qui vont dépendre de la valeur allouée au **M-support (**support maximal) et au **M-confiance (**confiance maximal). Une fois les règles générées, nous pourrions déjà les lire ensemble avec les pertes qu'il y a eu pendant la génération des règles.
- **Le module d'extraction des règles d'association** qui consiste à retirer seulement les règles qui sont pertinentes en fonction des valeurs attribuées à la confiance et du support proposé. Toutes les règles non pertinentes seront élaguées dans le module suivant.
- **Le module d'élagage de règles d'association** qui consiste à élaguer les règles d'association redondantes qui sont en dessous de la confiance et du support choisis. Il élimine également toutes les règles acycliques.
- **Le module de génération de règles intéressantes** qui consiste à ne retenir que les règles qui sont importantes et les items qui vont constituer ces règles seront retenus pour constituer la nouvelle base de données. **[4]**

Le schéma architectural du fonctionnement de ce logiciel est celui présenté à la **Figure 6.2**

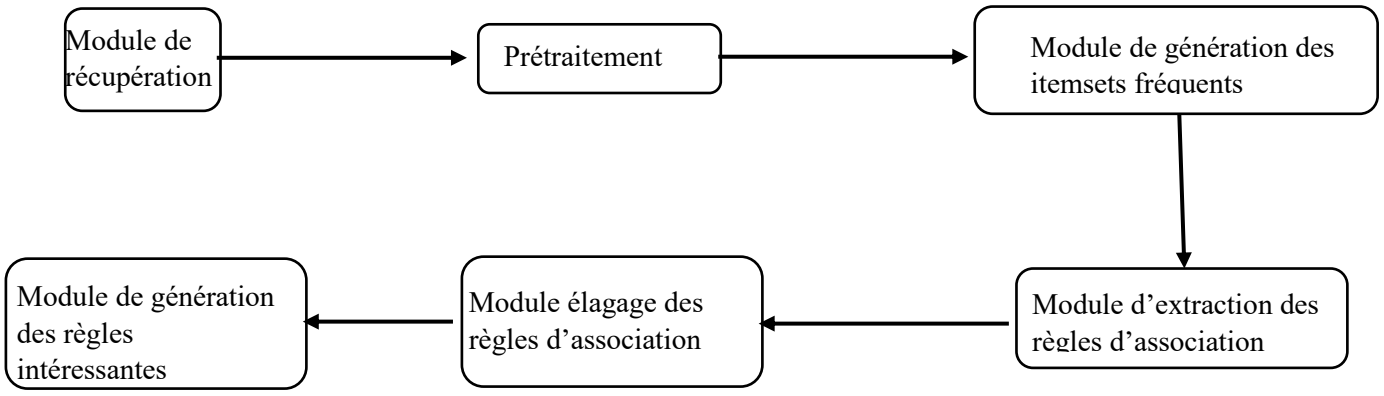

Figure 6.2. Schéma Architectural des modulaires du logiciel YamBob [4]

6.5.4.2.2. Modèle de fonctionnement

Le modèle de fonctionnement de notre logiciel est représenté par le schéma de principe ci-dessous. Ce schéma expliquer le fonctionnement détaillé du logiciel à utiliser pour faire les règles d'association.

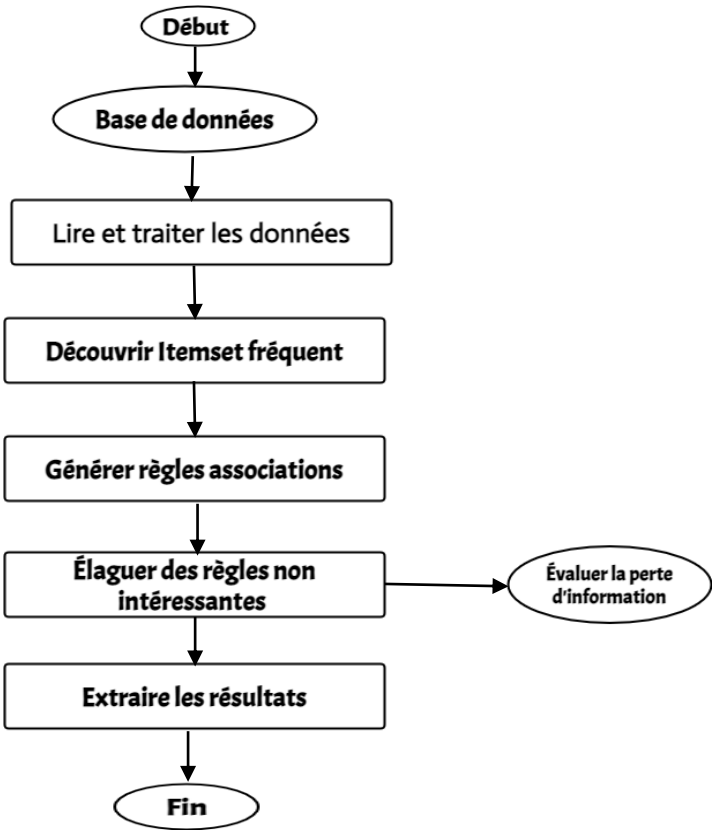

Figure 6.3. Schéma fonctionnel du principe d'extraction des règles d'association intéressantes [4]

6.5.4.2.3. Mesures d'évaluation

Il existe plusieurs mesures éligibles d'évaluation des règles d'association comme celles qui sont développées dans **[65]** et **[66],** mais pour ce qui concerne ce travail, nous allons retenir trois (3) mesures qui sont le support, la confiance et le lift.

- Le Lift sera utilisé comme indicateur de pertinence de nos règles et permettra de filtrer les règles. C'est grâce à lui que nous serons capables de mesurer le niveau d'amélioration que la règle d'association a apportée par rapport aux transactions faites.
- Le support sera utilisé pour vérifier la prémisse, la conséquence de la règle et enfin pour élaguer les règles non intéressantes **[66]** qui seront en dessous du seuil fixé. Ses choix seront situés de l'intervalle compris entre [0 et 1].
- La confiance est choisie pour établir le rapport d'entité qui vérifie le conséquent parmi celles qui vérifient la prémisse de la règle. **[66]** Sa valeur aussi sera comprise dans l'intervalle [0 à 1].

On sait par exemple que :

- sup(ATT) = P(ATT) représente le support de l'antécédent
- sup(TYPE) = P(TYPE) représente le support du conséquent
- sup(ATT→ TYPE) = supp(ATTTYPE) = P(ATTTYPE)
- conf(ATT →TYPE)= P(TYPE/ATT)
- et Lift(ATT→TYPE) = $\frac{\mathrm{P(TYPE/ATT)}}{\mathrm{P(TYPE)}}$

L'association ATT → TYPE peut-être considérée comme une règle intéressante que si :

1. supp (ATT ∪ TYPE) ≥ min_supp,
2. |supp (ATT ∪ TYPE) −supp (ATT) supp (TYPE) | ≥ min_interest,
3. conf (ATT → TYPE) = supp (ATT ∪ TYPE) / supp (ATT) ≥ min_conf **[67]**
4. lift (ATT → TYPE) > 1

Dans notre cas, nous avons choisi les valeurs ci-après présentées au **Tableau 6.3**:

Nous avons effectué le choix des valeurs pour chacune des mesures de façon subjective jusqu'à obtenir des bons résultats avec des règles intéressantes de la forme (x→y). Un total de trois (3) expérimentations ont été fait avec les valeurs ci-après :

Support	Confiance	Lift
0,08	0,8	1
0,12	0,7	1
0,008	0,9	1

Tableau 6.3. Valeurs des métriques

6.5.4.2.4. Évaluation des résultats

L'analyse faite avec les mesures ci-dessus indiquées nous donne les cinq (5) items suivants les plus fréquents :

- Idle Max() avec 271
- Idle Mean() avec 271
- Idle Min() avec 271
- ATT() avec 188
- Et TYPE() avec 188

Quant aux règles d'association, nous avons obtenu des règles qui varie selon les valeurs des métriques dont les détails seront commentés dans le chapitre suivant. Et comme le logiciel a un module d'élagage de règles d'association, toutes celles qui sont redondantes et non intéressantes ont été élaguées au point de rester avec un nombre de règles plus petit que le nombre de départ.

Pour notre cas, nous avons sélectionné seulement les règles qui sont en association avec la variable ou l'item ATT, car il s'agissait d'optimiser le système de détection d'intrusions en confirmant les vraies attaques et nous avons vu le nombre de règles diminuées.

Dans chacun des cas expérimentés, nous avons retenu les règles selon les critères ci-dessous :

1. Celles dont l'antécédent(lhs) ou le conséquent(rhs) est un ensemble des items puis ayant une relation avec la variable ATT
2. Celles dont le M-support est supérieur au seuil défini,
3. Celles dont la confiance est égale à 1,
4. Celles dont le lift est supérieur ou égal à 1,
5. Plus la valeur du support, de la confiance et du lift augmente (largement au-delà du seuil), plus la règle devient intéressante.

6.5.4.2.5. Variables à retenir

Après l'exécution des règles d'association, nous allons retenir seulement les variables qui ont des relations intéressantes avec la variable attaque (y compris elle-même). Cela va nous permettre de redimensionner notre dataset. Toutes ses variables qui ne sont pas dans une des règles qui respectent les critères définis sont élaguées de notre dataset.

Ainsi, notre dataset a été réduit de 14 à 5 variables ci-dessous:

1. FwdIATStd
2. ATT
3. FwdIATMax
4. FwdIATTot
5. FwdIATMean

Les items FlowIATMin, FwdPkts, BwdPkts, FwdIATMin, TYPE, FlowIATStd, IdleMin, IdleMax et IdleMean seront supprimés de notre dataset, car leurs relations avec l'item ATT ne sont pas intéressantes.

6.5.4.3. Analyse basée sur la régression logistique

Pour optimiser notre système de détection d'intrusion, nous avons besoin d'un modèle statistique de prédiction qui sera obtenu avec la régression logistique. Grâce à ce modèle, nous pourrions estimer la présence d'une attaque ou pas dans le réseau véhiculaire V2G.

Nous allons utiliser la régression logistique binaire, car les valeurs de notre variable à prédire ou à expliquer (ATT) sont binaires soit 0 quand il n'y a pas d'attaque et 1 quand il y a attaque. Dans notre cas, nous avons NON pour traduire le manque d'attaque et OUI pour confirmer qu'il y a attaque.

6.5.4.3.1. Description du modèle logistique binaire à utiliser

Comme la régression logistique vise à maximiser le log-vraisemblance et la marge **[68],** dans notre cas, ATT est considérée comme variable dépendante (Y) où les valeurs de la variable dichotomique ne peuvent avoir que deux (2) valeurs : 0 ou 1 (NON ou OUI) alors que les autres variables sont considérées comme indépendantes (X) et peuvent avoir des valeurs discrètes ou continues.

Pour modéliser, nous allons utiliser le dataset redimensionné après l'exécution des règles d'association maximale, qui a été présenté précédemment et qui dispose de cinq (5) variables qui sont utiles dans le modèle (variables de l'équation).

6.5.4.3.2. Présentation du dataset

Notre dataset que contient 5 variables, dont une (1) dite dépendante et quatre (4) indépendantes, avec chacune 283 observations sans une seule donnée manquante. La variable ATT qui n'a que deux (2) possibilités : soit il y a une attaque ou soit il n'y a pas d'attaques à 188 observations qui traduisent le manque d'attaque (soit 66,4%) et 95 observations qui traduisent la présence d'attaque (soit 33,6%). Le manque d'attaque a reçu le code NON et la présence de l'attaque le code OUI.

6.5.4.3.3. Modélisation

Dans cette partie de notre travail, nous allons ressortir une équation de la régression logistique avec nos variables significativement fortes qui peuvent nous amener à confirmer qu'il y a eu attaque ou non dans le réseau V2G.

6.5.4.3.3.1. Paramétrage et variables de l'équation

Le logiciel R est celui qui sera utilisé dans notre cas avec les configurations ci-après :

- Intervalle de confiance 95%
- P-valu de 0,05
- Y= ATT
- X1= FwdIATStd
- X2= FwdIATMax
- X3= FwdIATTot
- X4= FwdIATMean

Le modèle déterminera les chances (apostériori) qu'une variable (parmi les 4 qui sont indépendantes) prise au hasard est susceptible de confirmer une attaque par rapport à ses valeurs.

Les résultats obtenus qui seront commentés dans le chapitre suivant nous donnent une forte corrélation des variables entre elles et nous donnent un modèle hautement significatif au seuil de 5%, car les tests des coefficients du modèle sont élevés.

Étant donné que nous sommes dans la régression logistique binaire où Y représente la valeur de la variable dépendante dichotomique qui ne peut prendre que la valeur NON pour présenter le manque d'attaque, ou la valeur OUI pour présenter la présence d'une attaque donc le succès. Tandis que X représente les valeurs des différents attributs prédictifs relatifs à chaque échantillon ou participant pouvant avoir des valeurs discrètes ou continues. **[43]**

Le modèle aura une formule de la régression logistique sous la forme de celle présenter dans le chapitre de l'état de l'art.

6.5.4.3.3.2. Équation du modèle

Le modèle que nous avons développé améliore la classification des IDS dans le réseau véhiculaire V2G et sa matrice de confusion qui varie selon les cas est de presque 72%.

Dans l'ensemble des données utilisées dans le modèle, 70% sont des données d'entrainement utilisées pour entrainer notre algorithme et 30% des données sont les données de test que nous avons utilisé pour vérifier la performance des résultats.

L'équation finale du modèle est :

$$\mathbb{P}(Y=1|X)=\frac{e^{\beta 0+\beta 1X1+\beta 2X2+\beta 3X3+\beta 4X4}}{1+e^{\beta 0+\beta 1X1+\beta 2X2+\beta 3X3+\beta}}$$

Avec :

P(Y=1|X) = la probabilité d'attaque

Y = variable à expliquer (ATT) ou attaque à prédire

X= variables explicatives ou indépendantes

6.6. Conclusion

Plusieurs changements ont été apportés dans ce système qui vient améliorer le système de détection d'intrusion dans le réseau véhiculaire V2G. La méthodologie utilisée dans notre solution combine la corrélation de Pearson, les règles d'association et la régression logistique binaire dont chacune d'elle diminue la possibilité de signaler une fausse attaque. À la fin de la solution, un modèle est proposé avec seulement les variables susceptibles de confirmer une attaque.

Notre solution est capable d'archiver les fausses attaques et de dénoncer les vraies attaques.

CHAPITRE 7 : SIMULATIONS ET ANALYSE DES RÉSULTATS

Introduction

Dans ce chapitre, nous allons présenter et commenter les résultats des différentes étapes de notre approche. C'est les cas de la corrélation de Pearson, l'extraction des règles d'association et de la régression logistique binaire qui ont été utilisées chacune avec son environnement logiciel.

7.1. Données

Le dataset qui a été utilisé dans ce mémoire est celui qui a été généré dans MiniV2G avec l'ensemble des outils que nous avons présenté dans le chapitre précèdent. Dans ce dataset, deux (2) scénarios des IDS ont été exploités : Homme du milieu et le Déni de service comme le montre l'extrait dans le tableau **7.1**.

Ce dataset a 283 observations dont : 188 pour le sans attaque et 95 avec attaque (41 attaques avec le MIM et 54 avec le DOS).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	Flow Dur	Bwd Pkt	Fwd Pkts/s	Flow IAT Min	Flow IAT Mean	Flow Pkts/s	Flow IAT Max	Tot Fwd Pkts	Flow IAT Std	Fwd IAT Tot	Fwd IAT Mean	Fwd IAT Std	Fwd IAT Max	Fwd IAT Min	Subflow Fwd	Active Mean	Active Std	Active Max	Active Min	Idle Mean	Idle Std	Idle Max	Idle Min	ATT	type
2	817221	1.223659	1.2236592060164877	817221.0	817221.0	2.4473184120	817221.0	1	0.0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	O
3	497334	2.010721	4.021442330506259	213942.0	248667.0	6.0321634957	283392.0	2	49108.5659534	213942.0	213942.0	0.0	213942.0	213942.0	2	0	0	0	0	0	0	0	0	0	O
4	194909	5.130599	5.130599407928829	194909.0	194909.0	10.261198815	194909.0	1	0.0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	O
5	470937	2.123426	212.34262757014207	3026.0	4709.3699999999	214.46605384	21201.0	100	2850.55069163	465516.0	4702.1818181818	2864.1463320	21201.0	3026.0	100	0	0	0	0	0	0	0	0	0	O
6	545543	1.833036	3.6660721519660227	223913.0	272771.5	5.4991082279	321630.0	2	69096.3533372	223913.0	223913.0	0.0	223913.0	223913.0	2	0	0	0	0	0	0	0	0	0	O
7	1562584	0.639965	131.19294706716568	-253309.0	7622.3609756097	131.83291266	543574.0	205	47655.0874783	1562578.0	7659.696078431	47769.314750	543574.0	-253309.0	205	0	0	0	0	0	0	0	0	1	MIM
8	690189	1.448878	10.142149469203362	-12011.0	98598.428571428	11.591027964	363446.0	7	165102.120395	690196.0	115032.6666666	174475.62885	363446.0	-12011.0	7	0	0	0	0	0	0	0	0	1	MIM
9	1395538	0.716569	2.1497085711747013	-2.0	465179.33333333	2.8662780948	1395542.0	3	805717.704065	1395540.0	697770.0	986798.62584	1395542.0	-2.0	3	0	0	0	0	0	0	0	0	1	MIM
10	5	199999.9	199999.99999999997	5.0	5.0	399999.99999	5.0	1	0.0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	MIM
11	970414	1.030488	207.12809172167755	-250843.0	4827.9303482587	208.15857974	45541.0	201	29512.8110734	970431.0	4852.155000000	29584.867513	45541.0	-250843.0	201	0	0	0	0	0	0	0	0	1	MIM
12	1649884	0.606103	1.818309650860303	-3.0	549961.33333333	2.4244128678	1649883.0	3	952560.105512	1649887.0	824943.5	1166640.6290	1649883.0	4.0	3	0	0	0	0	0	0	0	0	1	DOS
13	69	14492.75	14492.753623188406	69.0	69.0	28985.507246	69.0	1	0.0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	DOS
14	1109355	0.901424	6.309972912187712	-22981.0	158479.28571428	7.2113976139	679133.0	7	274721.020775	1109347.0	184891.1666666	291043.35154	679133.0	-22981.0	7	0	0	0	0	0	0	0	0	1	DOS
15	7	142857.1	142857.14285714287	7.0	7.0	285714.28571	7.0	1	0.0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	DOS
16	-10	0.0	0.0	-10.0	-10.0	-199999.9999	-10.0	1	0.0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	DOS

Tableau 7.1. Exemple d'un extrait du dataset

7.1.1. Simulation du dataset

Comme expliquer dans le chapitre précédent, deux scénarios ont été simulés pour obtenir notre dataset : l'homme du milieu et le déni de service.

7.1.1.1. Environnement de simulation

Les applications utilisées dans la simulation de notre système ne sont compatibles que sur l'environnement du système d'exploitation Linux (distribution Ubuntu) de la version 12.10. Pour ce faire, une machine virtuelle a été installée avec le logiciel de virtualisation multiplateforme Oracle VM VirtualBox. C'est un des logiciels les plus populaires à cause de son code source qui est ouvert, mais aussi, il dispose de plusieurs fonctionnalités : il est doté des grandes performances et peut fonctionner dans plusieurs systèmes d'exploitation (Windows, Linux, Macintosh et solaris).

Dans cet environnement virtuel, on a été installé Ubuntu (la version 12.10) dans lequel était installé le MiniV2G qui est venu avec sa suite d'applications. L'ordinateur virtuel a les caractéristiques suivantes :

Operating system	Linux (Ubuntu 12.10) 64 bits
Processors	3
Processor frequency	3,6 Ghz
Base memory	6060 MB
Video memory	16 MB
Graphic controller	VMSVGA
Hard disk	SATA 32 GB
Network adapter	Intel PRO/1000MT

Tableau 7.2. Caractéristiques de l'ordinateur utilisé dans la simulation

7.1.1.2. Paramétrages

Sur cet ordinateur est installé un réseau V2G qui a généré le datraset après 6 minutes comme le montre le **Tableau 7.3.**

Véhicule électrique	6
Borne	3
Nombre de simulation	5
Durée de l'opération de simulation	6 minutes

Tableau 7.3. Paramétrage de la simulation

Dans l'ensemble, un total de cinq (5) simulations ont été réalisées afin d'obtenir un résultat fiable sur 6 véhicules électriques et 3 bornes de recharges. Les scénarios utilisés étaient basés sur les facteurs ci-après, dont les algorithmes pour les attaques positives ont été présentés dans le chapitre précédent.

Il s'agit :

- Du scénario sans attaque ;
- Du scénario avec attaque MIM (homme du milieu) ;
- Du scénario avec attaque déni de service.

Concernant les métriques retenues, nous avons utilisé celles exerçant une influence sur la performance du réseau et ce sont les 24 qui se trouvent dans notre dataset dont la description a déjà été faite dans le chapitre précédent.

7.1.1.3. Scénarios

Nous allons présenter les architectures des trois scénarios qui ont été utilisés.

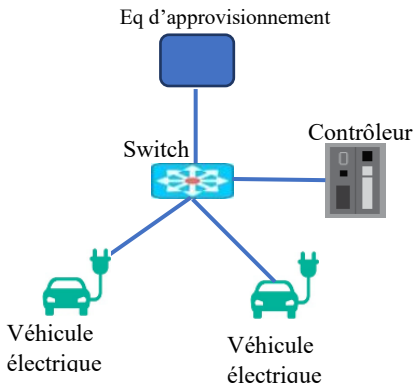

Figure 7.1.1. Scénario sans attaque

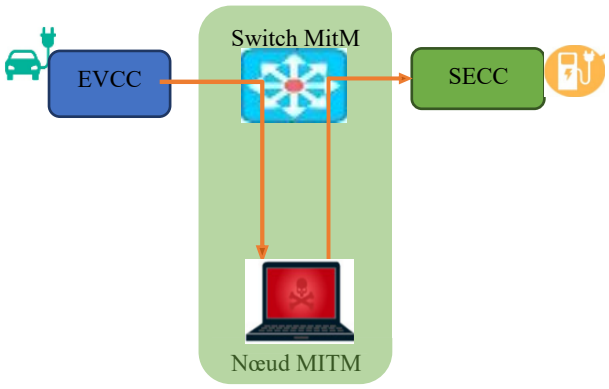

Figure 7.1.2. Scénario avec attaque de l'homme du milieu

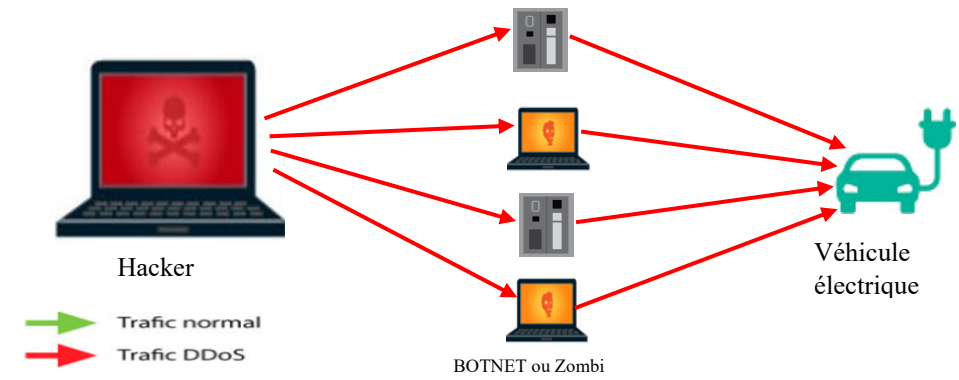

Figure 7.1.3. Scénario avec attaque de déni de service

7.1.1.3.1. Scénario sans attaque

Pour réaliser le scénario sans attaque de notre simulation, nous avons créé des interfaces aux niveaux de chaque équipement présent dans la **Figure 7.1.1** afin de chercher à capturer les différents paquets qui circulent dans le réseau et d'éditer leurs contenus. Cette technique a été faite avec l'outil Wireshark qui nous a donné la possibilité non seulement d'afficher les données, mais aussi de les enregistrer au format PCAP. Les résultats dans cette simulation ne montrent aucune anomalie, car il s'agit de zéro attaque.

7.1.1.3.2. Scénario avec attaque de type Homme du milieu

Quant au scénario avec attaque de type homme du milieu, nous assistons à une redirection par le commutateur des paquets venant de l'EVCC vers le nœud du hacker qui modifie puis transmet les paquets modifiés au SECC comme le montre la **Figure 7.1.2**. Il y a donc la présence d'un nœud supplémentaire entre le VECC et le SECC qui se connecte au réseau V2G via un commutateur et qui perturbe le fonctionnement du réseau. La communication ne se fait plus de façon directe entre le VECC et le SECC, elle passe par un hacker comme le montre la même **Figure 7.1.2** et c'est lui qui gère les communications en prenant une décision sur les paquets s'ils seront transférés, interrompus, modifiés ou s'il faut fabriquer un nouveau paquet à envoyer vers la destination.

Dans ce scénario, l'outil Wireshark a été utilisé pour sniffer les interfaces qui ont été créées dans les différentes entités de notre réseau V2G. Grâce à cette opération, les paquets circulant dans le réseau ont été écoutés et capturés pour être enregistrés sous le format PCAP.

7.1.1.3.3. Scénario avec attaque de type déni de service

Enfin pour le scénario du type DOS, L'hacker envoie plusieurs requêtes en continu vers des contrôleurs jusqu'à ce que le service du réseau qui est visé soit perturbé. Dans notre cas, nous avons utilisé le DOS distribué comme le montre la **Figure 7.1.3** qui s'exécute en deux phases ci-après :

1. Infection des Botnets : Le hacker commence par attaquer l'ensemble des terminaux qui seront utilisés pour inonder le VE cible. Ses terminaux portent le nom de Zombie ou Botnet.
2. Déclanchement du DOS : les terminaux Botnet vont à leur tour inonder le VE cible de façon synchrone comme le montre la **Figure 7.1.3**. Cette inondation va perturber le fonctionnement du service réseau ciblé qui sera ralenti jusqu'à le rendre parfois indisponible. Pour notre cas, le service visé est la recharge du véhicule électrique.

Comme dans les autres scénarios, les interfaces de chacune des entités ont été sniffées par l'outil Wireshark afin de produire des données stockées dans un fichier au format PCAP.

7.1.2. Création du dataset

Après avoir réalisé les simulations des trois (3) scénarios, nous nous retrouvons avec trois (3) fichiers au format PCAP créés à la suite de l'écoute et de l'enregistrement des interfaces des entités de notre réseau V2G, car pour chaque scénario, nous avions enregistré un fichier PCAP que nous avons nettoyé puis ajouté une colonne ATTAQUE, car l'objectif est non seulement d'obtenir un dataset pour les attaques, mais aussi parce que tous les scénarios réalisés concernent les attaques dans le réseau V2G simulé. Nous avons les deux cas ci-après :

1. ATTAQUE égale 0 pour les cas de scénario de simulation sans attaque,
2. ATTAQUE égale 1 pour les cas de scénario de simulation avec attaque tant MIM ou DOS, car dans ces deux cas, il y a la présence d'une attaque.

Nous avons ensuite ajouté une colonne types d'attaque et nous avons obtenu les trois cas ci-après :

1. Type d'attaque égal à 0 pour les scénarios sans attaque ;
2. Type d'attaque égal MIM pour les attaques du type homme du milieu ;
3. Type d'attaque égal DOS pour les attaques du type déni de service.

Nous avons enfin terminé notre simulation en fusionnant les trois datasets partiels, ce qui nous a donné un seul dataset du V2G que nous allons exploiter dans notre approche.

7.2. Corrélation

La corrélation ne tiendra compte que des données traitées en considérant la corrélation significative à 0,01 et a 0,05.

Matrice de corrélation

C'est une matrice va nous permettre de regrouper les différentes corrélations de plusieurs variables entre elles, les coefficients indiquant l'influence que les variables ont les unes sur les autres.

Corrélations[d]

Variables		Variables																								
		FlowDuration	BwdPktss	FwdPktss	FlowIATMin	FlowIATMean	FlowPktss	FlowIATMax	TotFwdPkts	FlowIATStd	FwdIATTot	FwdIATMean	FwdIATStd	FwdIATMax	FwdIATMin	SubflowFwd Pkts	ActiveMean	ActiveStd	ActiveMax	ActiveMin	IdleMean	IdleStd	IdleMax	IdleMin	TYPE
FlowDuration	Corrélation de Pearson	1	-0.077	-0.077	.130**	.790**	0.047	.991**	-0.074	.919**	.969**	.870**	.908**	.964**	-.143**	-0.074	.224**	.[c]	.224**	.224**	.816**	.[c]	.816**	.816**	0.109
	Sig. (bilatérale)		0.199	0.197	0.029	0.000	0.431	0.000	0.212	0.000	0.000	0.000	0.000	0.000	0.016	0.212	0.000		0.000	0.000	0.000		0.000	0.000	0.068
BwdPktss	Corrélation de Pearson	-0.077	1	1.000**	-0.032	-0.060	.643**	-0.067	-0.064	-0.048	-0.067	-0.043	-0.046	-0.059	-0.007	-0.064	-0.024	.[c]	-0.024	-0.024	-0.024	.[c]	-0.024	-0.024	.227**
	Sig. (bilatérale)	0.199		0.000	0.593	0.318	0.000	0.265	0.282	0.417	0.262	0.470	0.443	0.326	0.906	0.282	0.694		0.694	0.694	0.690		0.690	0.690	0.000
FwdPktss	Corrélation de Pearson	-0.077	1.000**	1	-0.032	-0.060	.643**	-0.067	-0.063	-0.049	-0.067	-0.043	-0.046	-0.059	-0.007	-0.063	-0.024	.[c]	-0.024	-0.024	-0.024	.[c]	-0.024	-0.024	.228**
	Sig. (bilatérale)	0.197	0.000		0.590	0.315	0.000	0.262	0.287	0.414	0.260	0.467	0.441	0.324	0.902	0.287	0.692		0.692	0.692	0.688		0.688	0.688	0.000
FlowIATMin	Corrélation de Pearson	.130**	-0.032	-0.032	1	.629**	0.020	.164**	-.144**	-0.077	-.117**	-0.063	-0.073	-0.098	0.092	-.144**	-0.042	.[c]	-0.042	-0.042	.234**	.[c]	.234**	.234**	-.181**
	Sig. (bilatérale)	0.029	0.593	0.590		0.000	0.742	0.006	0.015	0.197	0.050	0.290	0.222	0.101	0.124	0.015	0.477		0.477	0.477	0.000		0.000	0.000	0.002
FlowIATMean	Corrélation de Pearson	.790**	-0.060	-0.060	.629**	1	0.037	.834**	-.161**	.723**	.639**	.732**	.727**	.677**	0.019	-.161**	0.038	.[c]	0.038	0.038	.828**	.[c]	.828**	.828**	-0.005
	Sig. (bilatérale)	0.000	0.318	0.315	0.000		0.540	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.752	0.006	0.529		0.529	0.529	0.000		0.000	0.000	0.937
FlowPktss	Corrélation de Pearson	0.047	.643**	.643**	0.020	0.037	1	0.041	0.040	0.030	0.041	0.026	0.028	0.036	0.004	0.040	0.014	.[c]	0.014	0.014	0.015	.[c]	0.015	0.015	-0.042
	Sig. (bilatérale)	0.431	0.000	0.000	0.742	0.540		0.494	0.505	0.618	0.491	0.657	0.638	0.547	0.943	0.505	0.809		0.809	0.809	0.806		0.806	0.806	0.484
FlowIATMax	Corrélation de Pearson	.991**	-0.067	-0.067	.164**	.834**	0.041	1	-.136**	.944**	.953**	.906**	.936**	.966**	-0.094	-.136**	.210**	.[c]	.210**	.210**	.862**	.[c]	.862**	.862**	0.105
	Sig. (bilatérale)	0.000	0.265	0.262	0.006	0.000	0.494		0.022	0.000	0.000	0.000	0.000	0.000	0.114	0.022	0.000		0.000	0.000	0.000		0.000	0.000	0.078
TotFwdPkts	Corrélation de Pearson	-0.074	-0.064	-0.063	-.144**	-.161**	0.040	-.136**	1	-0.113	-0.046	-0.110	-0.108	-0.111	-.463**	1.000**	-0.047	.[c]	-0.047	-0.047	-0.061	.[c]	-0.061	-0.061	0.051
	Sig. (bilatérale)	0.212	0.282	0.287	0.015	0.006	0.505	0.022		0.058	0.441	0.066	0.069	0.063	0.000	0.000	0.429		0.429	0.429	0.306		0.306	0.306	0.389
FlowIATStd	Corrélation de Pearson	.919**	-0.048	-0.049	-0.077	.723**	0.030	.944**	-0.113	1	.943**	.989**	.999**	.974**	-0.042	-0.113	.133**	.[c]	.133**	.133**	.860**	.[c]	.860**	.860**	.143**
	Sig. (bilatérale)	0.000	0.417	0.414	0.197	0.000	0.618	0.000	0.058		0.000	0.000	0.000	0.000	0.486	0.058	0.025		0.025	0.025	0.000		0.000	0.000	0.016
FwdIATTot	Corrélation de Pearson	.969**	-0.067	-0.067	-.117**	.639**	0.041	.953**	-0.046	.943**	1	.891**	.931**	.991**	-.159**	-0.046	.233**	.[c]	.233**	.233**	.763**	.[c]	.763**	.763**	.154**
	Sig. (bilatérale)	0.000	0.262	0.260	0.050	0.000	0.491	0.000	0.441	0.000		0.000	0.000	0.000	0.007	0.441	0.000		0.000	0.000	0.000		0.000	0.000	0.010
FwdIATMean	Corrélation de Pearson	.870**	-0.043	-0.043	-0.063	.732**	0.026	.906**	-0.110	.989**	.891**	1	.993**	.933**	0.015	-0.110	0.055	.[c]	0.055	0.055	.865**	.[c]	.865**	.865**	.139**
	Sig. (bilatérale)	0.000	0.470	0.467	0.290	0.000	0.657	0.000	0.066	0.000	0.000		0.000	0.000	0.796	0.066	0.356		0.356	0.356	0.000		0.000	0.000	0.019
FwdIATStd	Corrélation de Pearson	.908**	-0.046	-0.046	-0.073	.727**	0.028	.936**	-0.108	.999**	.931**	.993**	1	.965**	-0.041	-0.108	0.114	.[c]	0.114	0.114	.868**	.[c]	.868**	.868**	.149**
	Sig. (bilatérale)	0.000	0.443	0.441	0.222	0.000	0.638	0.000	0.069	0.000	0.000	0.000		0.000	0.488	0.069	0.055		0.055	0.055	0.000		0.000	0.000	0.012
FwdIATMax	Corrélation de Pearson	.964**	-0.059	-0.059	-0.098	.677**	0.036	.966**	-0.111	.974**	.991**	.933**	.965**	1	-0.097	-0.111	.222**	.[c]	.222**	.222**	.809**	.[c]	.809**	.809**	.146**
	Sig. (bilatérale)	0.000	0.326	0.324	0.101	0.000	0.547	0.000	0.063	0.000	0.000	0.000	0.000		0.104	0.063	0.000		0.000	0.000	0.000		0.000	0.000	0.014
FwdIATMin	Corrélation de Pearson	-.143**	-0.007	-0.007	0.092	0.019	0.004	-0.094	-.463**	-0.042	-.159**	0.015	-0.041	-0.097	1	-.463**	-0.070	.[c]	-0.070	-0.070	-0.024	.[c]	-0.024	-0.024	-.360**
	Sig. (bilatérale)	0.016	0.906	0.902	0.124	0.752	0.943	0.114	0.000	0.486	0.007	0.796	0.488	0.104		0.000	0.242		0.242	0.242	0.692		0.692	0.692	0.000
SubflowFwdPkts	Corrélation de Pearson	-0.074	-0.064	-0.063	-.144**	-.161**	0.040	-.136**	1.000**	-0.113	-0.046	-0.110	-0.108	-0.111	-.463**	1	-0.047	.[c]	-0.047	-0.047	-0.061	.[c]	-0.061	-0.061	0.051
	Sig. (bilatérale)	0.212	0.282	0.287	0.015	0.006	0.505	0.022	0.000	0.058	0.441	0.066	0.069	0.063	0.000		0.429		0.429	0.429	0.306		0.306	0.306	0.389
ActiveMean	Corrélation de Pearson	.224**	-0.024	-0.024	-0.042	0.038	0.014	.210**	-0.047	.133**	.233**	0.055	0.114	.222**	-0.070	-0.047	1	.[c]	1.000**	1.000**	.286**	.[c]	.286**	.286**	-0.003
	Sig. (bilatérale)	0.000	0.694	0.692	0.477	0.529	0.809	0.000	0.429	0.025	0.000	0.356	0.055	0.000	0.242	0.429			0.000	0.000	0.000		0.000	0.000	0.966
ActiveStd	Corrélation de Pearson	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]
	Sig. (bilatérale)																									
ActiveMax	Corrélation de Pearson	.224**	-0.024	-0.024	-0.042	0.038	0.014	.210**	-0.047	.133**	.233**	0.055	0.114	.222**	-0.070	-0.047	1.000**	.[c]	1	1.000**	.286**	.[c]	.286**	.286**	-0.003
	Sig. (bilatérale)	0.000	0.694	0.692	0.477	0.529	0.809	0.000	0.429	0.025	0.000	0.356	0.055	0.000	0.242	0.429	0.000			0.000	0.000		0.000	0.000	0.966
ActiveMin	Corrélation de Pearson	.224**	-0.024	-0.024	-0.042	0.038	0.014	.210**	-0.047	.133**	.233**	0.055	0.114	.222**	-0.070	-0.047	1.000**	.[c]	1.000**	1	.286**	.[c]	.286**	.286**	-0.003
	Sig. (bilatérale)	0.000	0.694	0.692	0.477	0.529	0.809	0.000	0.429	0.025	0.000	0.356	0.055	0.000	0.242	0.429	0.000		0.000		0.000		0.000	0.000	0.966
IdleMean	Corrélation de Pearson	.816**	-0.024	-0.024	.234**	.828**	0.015	.862**	-0.061	.860**	.763**	.865**	.868**	.809**	-0.024	-0.061	.286**	.[c]	.286**	.286**	1	.[c]	1.000**	1.000**	.131**
	Sig. (bilatérale)	0.000	0.690	0.688	0.000	0.000	0.806	0.000	0.306	0.000	0.000	0.000	0.000	0.000	0.692	0.306	0.000		0.000	0.000			0.000	0.000	0.028
IdleStd	Corrélation de Pearson	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]	.[c]
	Sig. (bilatérale)																									
IdleMax	Corrélation de Pearson	.816**	-0.024	-0.024	.234**	.828**	0.015	.862**	-0.061	.860**	.763**	.865**	.868**	.809**	-0.024	-0.061	.286**	.[c]	.286**	.286**	1.000**	.[c]	1	1.000**	.131**
	Sig. (bilatérale)	0.000	0.690	0.688	0.000	0.000	0.806	0.000	0.306	0.000	0.000	0.000	0.000	0.000	0.692	0.306	0.000		0.000	0.000	0.000			0.000	0.028
IdleMin	Corrélation de Pearson	.816**	-0.024	-0.024	.234**	.828**	0.015	.862**	-0.061	.860**	.763**	.865**	.868**	.809**	-0.024	-0.061	.286**	.[c]	.286**	.286**	1.000**	.[c]	1.000**	1	.131**
	Sig. (bilatérale)	0.000	0.690	0.688	0.000	0.000	0.806	0.000	0.306	0.000	0.000	0.000	0.000	0.000	0.692	0.306	0.000		0.000	0.000	0.000		0.000		0.028
TYPE	Corrélation de Pearson	0.109	.227**	.228**	-.181**	-0.005	-0.042	0.105	0.051	.143**	.154**	.139**	.149**	.146**	-.360**	0.051	-0.003	.[c]	-0.003	-0.003	.131**	.[c]	.131**	.131**	1
	Sig. (bilatérale)	0.068	0.000	0.000	0.002	0.937	0.484	0.078	0.389	0.016	0.010	0.019	0.012	0.014	0.000	0.389	0.966		0.966	0.966	0.028		0.028	0.028	

*. La corrélation est significative au niveau 0,05 (bilatéral).

**. La corrélation est significative au niveau 0,01 (bilatéral).

c. Calcul impossible, car au moins une des variables est une constante.

d. Toute observation incomplète N=283

Tableau 7.4. Matrice de corrélation de Pearson

En s'inspirant du **Tableau 7.4**, les valeurs avec un ou deux astérisques (*) sont toutes différentes de 0 à un niveau de signification alpha=0,05 (pour un astérisque) ou 0,01 (pour 2 astérisques) que nous avons choisi. Les valeurs qui manquent d'astérisques représentent le manque de corrélation qui est causé soit par la présence des observations incomplètes ou soit par la présence des observations incorrectes.

En observant le tableau des résultats de la corrélation de Pearson et en croisant toutes les variables entre elles dans le but de détecter la présence de relation entre elles, nous constatons que sur la diagonale principale, chacune des variables est en parfaite corrélation avec elle-même. Dans chaque cellule (à l'exception de la relation d'une variable avec elle-même qui donne 1), il y a deux valeurs : la première indique la corrélation de Pearson et la deuxième indique la signification de la relation.

Chaque fois que nous avons une relation significative détectée par la valeur de la corrélation de Pearson, on regarde son P-valu s'il est conforme aux conditions énumérées.

Notre but ici est de détecter les variables qui sont en corrélation avec la variable Type d'attaque afin de réduire la taille de notre dataset.

Les 12 variables Bwd Pkts/s ; Fwd Pkts/s ; Flow IAT Min ; Flow IAT Std ; Fwd IAT Tot ; Fwd IAT Mean ; Fwd IAT Std ; Fwd IAT Max ; Fwd IAT Min ; Idle Mean ; Idle Max ; Idle Min ; sont en corrélation avec la variable **Type d'attaque**, car non seulement les valeurs obtenues traduisent la présence d'une corrélation (faible, moyenne, forte ou parfaite), mais aussi parce que leur P-valu influence sur la corrélation. Chaque fois que la valeur de la corrélation entre une variable et la variable Type attaque est significative, on le confirme par son P-valu.

7.3. Essai des résultats des règles d'association maximale

Les règles d'association maximale tout comme la régression logistique font partie des méthodologies de base de notre solution. Maintenant, nous allons expérimenter le processus qui nous permettra d'extraire les règles d'association les plus pertinentes et d'élaguer celles qui sont moins pertinentes afin d'obtenir les variables utiles qui sont exploitées dans la modélisation.

Dans le chapitre précédent, nous avons dit que le choix des valeurs utilisées dans nos différents essais était fait de façon subjective jusqu'à obtenir les meilleurs résultats. Notre échantillon d'essai va se faire en trois (3) expérimentations comme nous l'avons montré dans le **Tableau 7.5.**

Support	Confiance	Lift
0,08	0,8	1
0,12	0,7	1
0,008	0,9	1

Tableau 7.5. Valeurs des métriques

7.3.1. Premier essai

Concernant notre premier essai, nous avons attribué aux métriques les valeurs suivantes :

- Support 0.08
- Confiance 0.9
- Lift 1

En utilisant ces valeurs, nous avons obtenu le résumé ci-après :

```
transactions as itemMatrix in sparse format with
 284 rows (elements/itemsets/transactions) and
 2043 columns (items) and a density of 0.007342144

most frequent items:
 IdleMax0 IdleMean0  IdleMin0      ATT0     TYPE0   (Other)
      271       271       271       188       188      3071

element (itemset/transaction) length distribution:
sizes
 15
284

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     15      15      15      15      15      15

includes extended item information - examples:
  labels
1    ATT
2   ATT0
3   ATT1
```

Figure 7.2. Résumé des données avec les valeurs de l'essai 1

Il ressort de ce résumé que la densité des transactions est de 284 lignes et 2043 colonnes avec cinq (5) items fréquents, dont IdleMax, IdleMean, IdleMin, ATT, TYPE et une statistique des données du dataset téléversé dans l'application. Les items les plus fréquents sont visualisés dans la **Figure 7.3** ci-après :

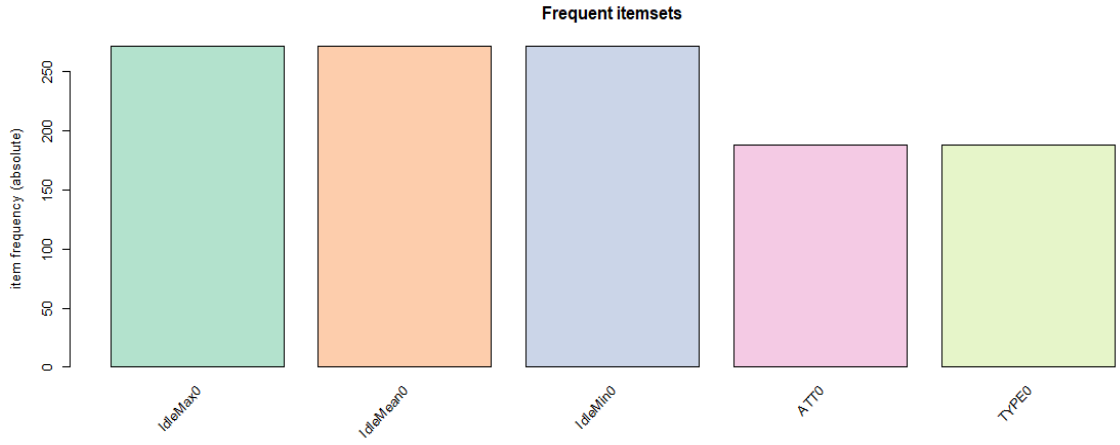

Figure 7.3. Items les plus fréquents de l'essai 1

Il ressort clairement de cette figure que les items ou variables Idle Max (), Idle Mean (), Idle Min () sont les plus fréquents avec 271 puis viennent ATT et TYPE avec chacune 188.

La structure de fonctionnement du logiciel YamBob **[4]** donne aussi la possibilité d'afficher les dix (10) premières règles d'association comme le montre le graphe de la **Figure 7.4.** Pour l'ensemble de ses règles, nous ne trouvons pas de règles intéressantes dans ce premier essai.

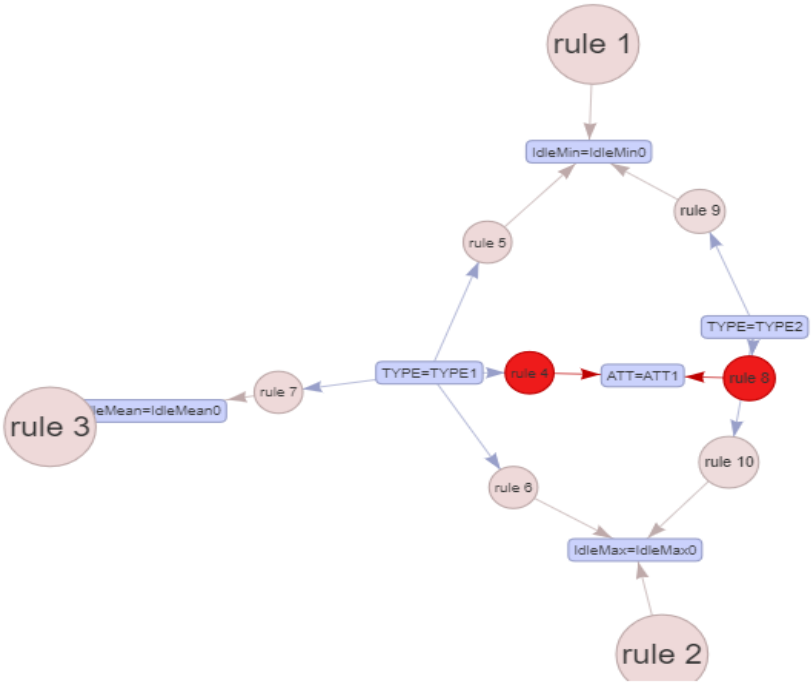

Figure 7.4. Graphe des 10 premières règles d'association avant élagage

Dans l'ensemble, avec les valeurs retenues dans notre premier essai, nous avons un total de 12.357 règles, mais après élagage, nous restons avec 88 règles d'association comme le montre la **Figure 7.5.**

	lhs	rhs	support	confidence	coverage	lift	count
1	{TYPE=TYPE1}	{ATT=ATT1}	0.141342756183746	1	0.141342756183746	2.97894736842105	40
2	{TYPE=TYPE2}	{ATT=ATT1}	0.19434628975265	1	0.19434628975265	2.97894736842105	55
3	{FlowIATStd=FlowIATStd0.0}	{FwdIATTot=FwdIATTot0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
4	{FwdIATTot=FwdIATTot0}	{FlowIATStd=FlowIATStd0.0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
5	{FlowIATStd=FlowIATStd0.0}	{FwdIATMean=FwdIATMean0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
6	{FwdIATMean=FwdIATMean0}	{FlowIATStd=FlowIATStd0.0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
7	{FlowIATStd=FlowIATStd0.0}	{FwdIATStd=FwdIATStd0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
8	{FwdIATStd=FwdIATStd0}	{FlowIATStd=FlowIATStd0.0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
9	{FlowIATStd=FlowIATStd0.0}	{FwdIATMax=FwdIATMax0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
10	{FwdIATMax=FwdIATMax0}	{FlowIATStd=FlowIATStd0.0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120

Showing 1 to 10 of 88 entries Previous 1 2 3 4 5 … 9 Next

Figure 7.5. Règles d'association restante après élagage

En appliquant les valeurs choisis du Support = 0,08; Confiance 0,9 et le Lift=1, nous ne trouvons pas de relations pertinentes entre deux variables ou items différents en dehors de quelques items qui donnent une forte relation avec elle-même donnant une confiance de 100%. Cette confiance de 100% est présente dans la plupart des relations à l'exception de quelques-unes qui sont supérieures à 50%.

Nous terminons notre premier essai en analysant la perte des informations avant et après l'élagage des règles non pertinentes comme le montrent les **Figures 7.6.1 ; 7.6.2 et 7.6.3**

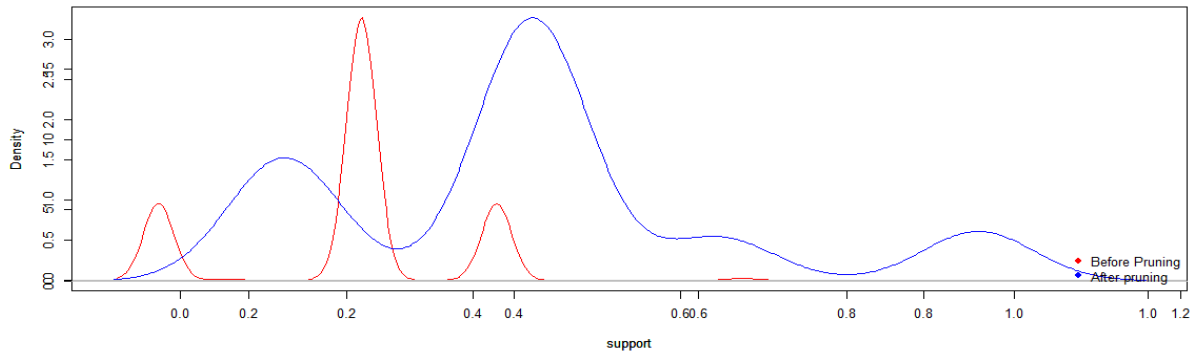

Figure 7.6.1. Densité de l'indice du support de l'essai 1

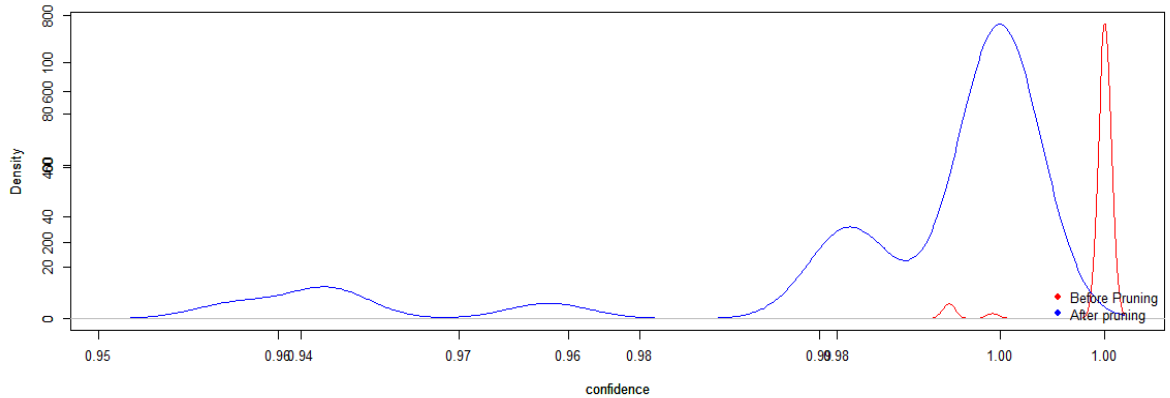

Figure 7.6.2. Densité de l'indice de la Confiance de l'essai 1

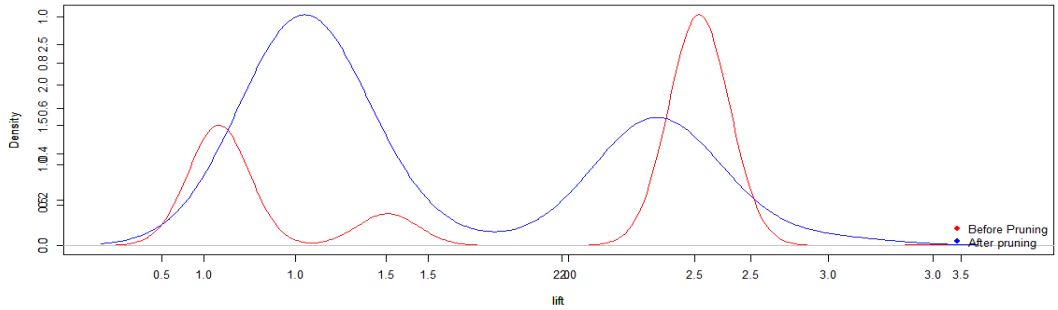

Figure 7.6.3. Densité de l'indice du Lift de l'essai 1

En faisant une lecture sur les trois (3) graphiques qui représentent chacune des métriques, nous voyons qu'il y a une perte importante des données entre l'avant et l'après l'élagage des règles. Avec la métrique support par exemple, la densité est très élevée au support égal à 0,2 avant l'élagage tandis qu'elle est plus élevée au support égal à 0.4 après élagage. Cela explique qu'il n'y a pas des règles pertinentes dans cet essai. Bien que la métrique confiance présente peu de perte jusqu'à 0,95; mais cela ne suffit pas pour confirmer la présence des règles pertinentes, car il faut tenir compte des autres métriques comme le support et le lift.

7.3.2. Deuxième essai

Concernant notre deuxième essai, nous avons attribué aux métriques les valeurs suivantes :

- Support 0.12
- Confiance 0.7
- Lift 1

En utilisant ces valeurs dans chacune des métriques, nous avons obtenu le même résumé que l'essai 1comme le montre la **Figure 7.7** ci-après :

```
transactions as itemMatrix in sparse format with
 284 rows (elements/itemsets/transactions) and
 2043 columns (items) and a density of 0.007342144

most frequent items:
 IdleMax0 IdleMean0  IdleMin0      ATT0     TYPE0   (Other)
      271       271       271       188       188      3071

element (itemset/transaction) length distribution:
sizes
 15
284

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     15      15      15      15      15      15

includes extended item information - examples:
  labels
1    ATT
2   ATT0
3   ATT1
```

Figure 7.7. Résumé des données avec les valeurs de l'essai 2

Il ressort de ce résumé que la densité des transactions reste le même que dans le premier essai, soit 284 lignes et 2043 colonnes avec cinq (5) items fréquents, dont IdleMax, IdleMean, IdleMin, ATT, TYPE et une statistique des données du dataset téléversé dans l'application. Les items les plus fréquents sont restés aussi les mêmes comme on peut le voir dans la **Figure 7.8** ci-après :

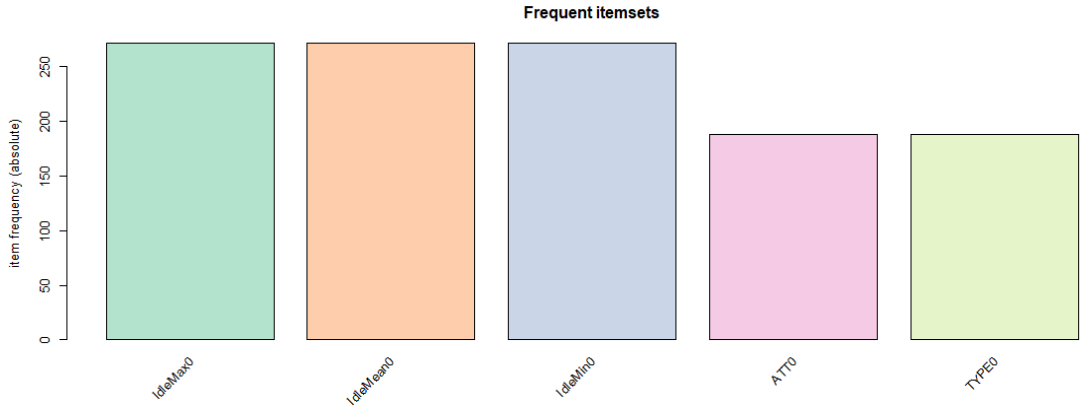

Figure 7.8. Items les plus fréquents de l'essai 2

Il ressort clairement de cette figure que les items ou variables Idle Max (), Idle Mean (), Idle Min () sont les plus fréquents avec 271 puis viennent ATT et TYPE avec chacune 188.

La structure de fonctionnement du logiciel YamBob **[4]** donne aussi la possibilité d'afficher les dix (10) premières règles d'association comme le montre le graphe de la **Figure 7.9**. Pour l'ensemble de ses règles, nous trouvons quelques règles intéressantes dans ce deuxième essai, mais qui ne sont pas trop pertinentes.

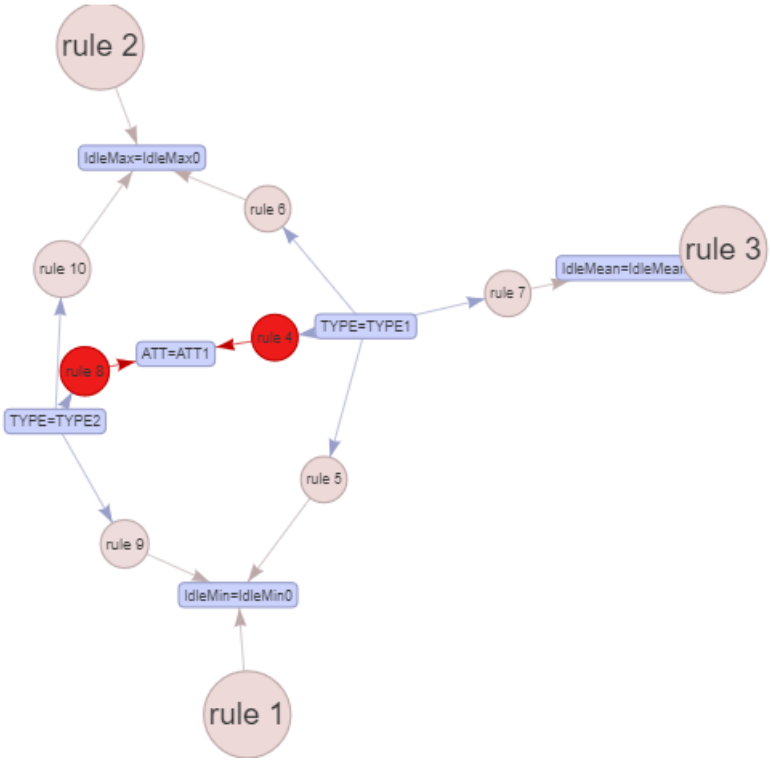

Figure 7.9. Graphe des 10 premières règles d'association avant élagage de l'essai 2

Dans l'ensemble, avec les valeurs retenues dans notre deuxième essai, nous avons un total de 13.365 règles, mais après élagage, nous restons avec 100 règles d'association comme le montre la **Figure 7.10**.

	lhs	rhs	support	confidence	coverage	lift	count
1	{TYPE=TYPE1}	{ATT=ATT1}	0.141342756183746	1	0.141342756183746	2.97894736842105	40
2	{TYPE=TYPE2}	{ATT=ATT1}	0.19434628975265	1	0.19434628975265	2.97894736842105	55
3	{FlowIATStd=FlowIATStd0.0}	{FwdIATTot=FwdIATTot0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
4	{FwdIATTot=FwdIATTot0}	{FlowIATStd=FlowIATStd0.0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
13	{FwdIATTot=FwdIATTot0}	{FwdIATMean=FwdIATMean0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
14	{FwdIATMean=FwdIATMean0}	{FwdIATTot=FwdIATTot0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
15	{FwdIATTot=FwdIATTot0}	{FwdIATStd=FwdIATStd0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
16	{FwdIATStd=FwdIATStd0}	{FwdIATTot=FwdIATTot0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
17	{FwdIATTot=FwdIATTot0}	{FwdIATMax=FwdIATMax0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120
18	{FwdIATMax=FwdIATMax0}	{FwdIATTot=FwdIATTot0}	0.424028268551237	1	0.424028268551237	2.35833333333333	120

Showing 1 to 10 of 45 entries (filtered from 100 total entries) Previous 1 2 3 4 5 Next

Figure 7.10. Règles d'association restante après élagage de l'essai 2

En appliquant les valeurs choisis du Support = 0,12; Confiance 0,7 et le Lift=1, nous ne trouvons pas des relations pertinentes entre deux variables ou items différents en dehors de quelques items qui donnent une forte relation avec elle-même donnant une confiance de 100%. Cette confiance de 100% est présente dans la plupart des relations à l'exception de quelques-unes.

Nous terminons notre deuxième essai en analysant la perte des informations avant et après l'élagage des règles non pertinentes comme le montrent les **Figures 7.11.1 ; 7.11.2 et 7.11.3**

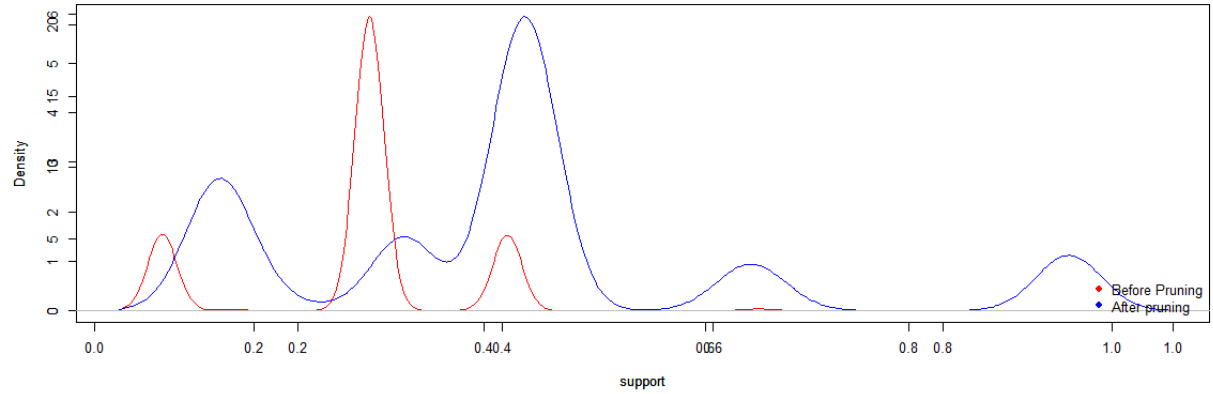

Figure 7.11.1. Densité de l'indice du support de l'essai 2

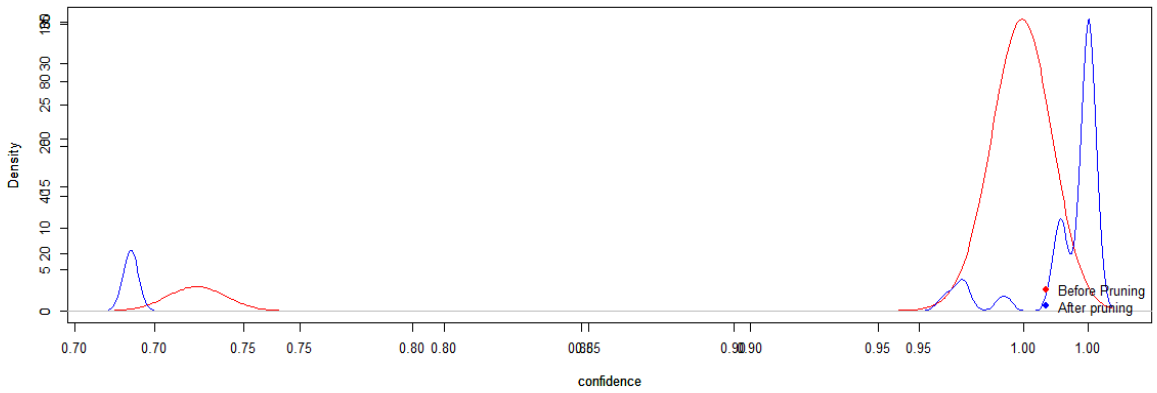

Figure 7.11.2. Densité de l'indice de la Confiance de l'essai 2

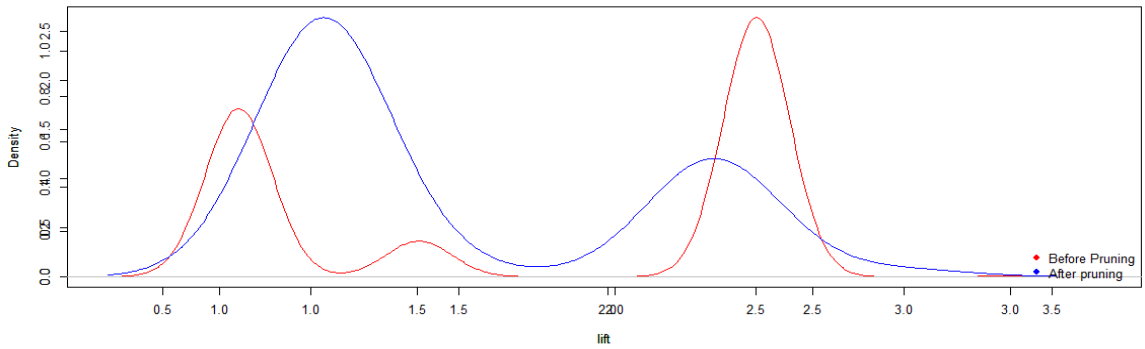

Figure 7.11.3. Densité de l'indice du Lift de l'essai 2

En faisant une lecture sur les trois (3) graphiques qui représentent chacune des métriques, nous voyons qu'il y a une perte importante des données entre l'avant et l'après l'élagage des règles. Les densités des indices sont très différentes sur toutes les métriques avant et après l'élagage. Cela explique qu'il n'y a pas des règles pertinentes dans cet essai.

À la différence avec le premier essai, les courbes des graphes du deuxième essai présente peu d'écart à certains niveaux. Cela se justifie qu'il y a quelques règles qui sont importantes, mais qui ne présentent pas une pertinence capable d'influencer les résultats.

7.3.3. Troisième essai

Concernant notre troisième et dernier essai, nous avons attribué aux métriques les valeurs suivantes :

- **M-Support** : 0,008
- **M-Confiance** : 0,9
- **lift** : > 1

En utilisant ces valeurs, nous avons obtenu le même résumé comme ceux du premier et deuxième essai en référence de la figure ci-après :

```
transactions as itemMatrix in sparse format with
 284 rows (elements/itemsets/transactions) and
 2043 columns (items) and a density of 0.007342144

most frequent items:
 IdleMax0 IdleMean0  IdleMin0      ATT0     TYPE0   (Other)
      271       271       271       188       188      3071

element (itemset/transaction) length distribution:
sizes
 15
284

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     15      15      15      15      15      15

includes extended item information - examples:
  labels
1    ATT
2   ATT0
3   ATT1
```

Figure 7.12. Résumé des données avec les valeurs du troisième essai

Il ressort de ce résumé que la densité des transactions est de 284 lignes et 2043 colonnes avec cinq (5) items fréquents, dont IdleMax, IdleMean, IdleMin, ATT, TYPE et une statistique des données du dataset téléversé dans l'application. Les items les plus fréquents sont les mêmes que dans les autres essais comme on peut le voir à la **Figure 7.13** ci-après :

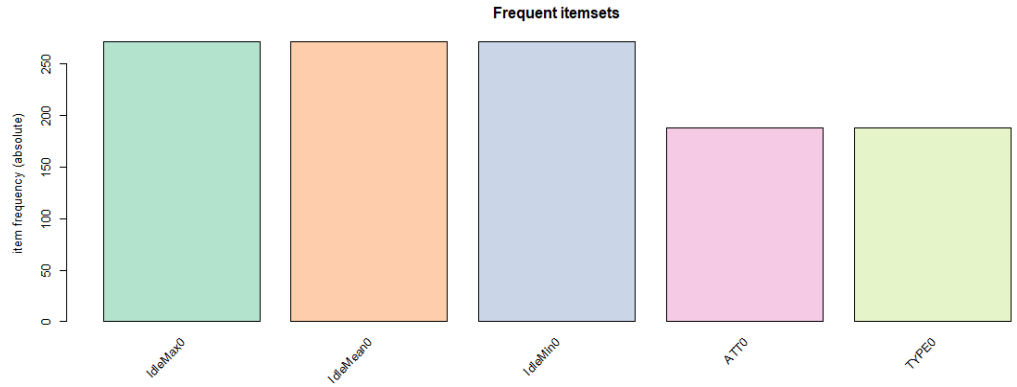

Figure 7.13. Items les plus fréquents de l'essai 3

Il ressort clairement de cette figure que les items ou variables Idle Max (), Idle Mean (), Idle Min () sont les plus fréquents avec 271 puis viennent ATT et TYPE avec chacune 188.

La structure de fonctionnement du logiciel YamBob **[62]** donne aussi la possibilité d'afficher les dix (10) premières règles d'association comme le montre le graphe de la **Figure 7.14**. Pour l'ensemble de ses règles, nous trouvons plusieurs règles intéressantes dans ce troisième essai.

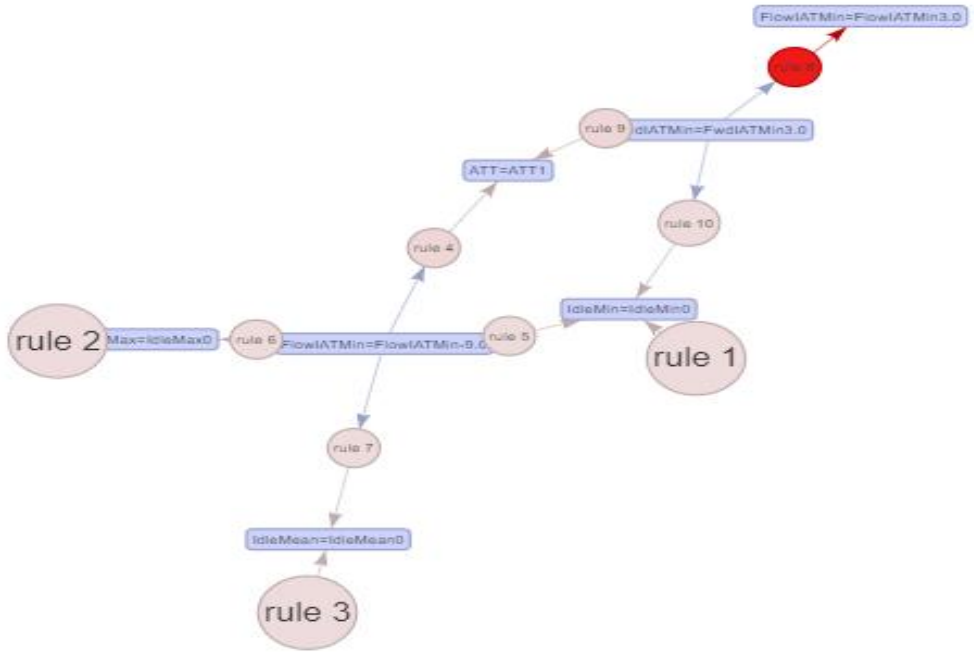

Figure 7.14. Graphe des 10 premières règles d'association avant élagage de l'essai 3

Dans l'ensemble, avec les valeurs retenues dans notre premier essai, nous avons un total de 234.056 règles, mais après élagage, nous restons avec 249 règles d'association comme le montre la **Figure 7.15.**

	lhs	rhs	support	confidence	coverage	lift	count
1	{}	{IdleMin=IdleMin0}	0.957597173144876	0.957597173144876	1	1	271
2	{}	{IdleMax=IdleMax0}	0.957597173144876	0.957597173144876	1	1	271
3	{}	{IdleMean=IdleMean0}	0.957597173144876	0.957597173144876	1	1	271
4	{FlowIATMin=FlowIATMin-9.0}	{ATT=ATT1}	0.0106007067137809	1	0.0106007067137809	2.97894736842105	3
5	{FlowIATMin=FlowIATMin-9.0}	{IdleMin=IdleMin0}	0.0106007067137809	1	0.0106007067137809	1.04428044280443	3
6	{FlowIATMin=FlowIATMin-9.0}	{IdleMax=IdleMax0}	0.0106007067137809	1	0.0106007067137809	1.04428044280443	3
7	{FlowIATMin=FlowIATMin-9.0}	{IdleMean=IdleMean0}	0.0106007067137809	1	0.0106007067137809	1.04428044280443	3
8	{FwdIATMin=FwdIATMin3.0}	{FlowIATMin=FlowIATMin3.0}	0.0106007067137809	1	0.0106007067137809	56.6	3
9	{FwdIATMin=FwdIATMin3.0}	{ATT=ATT1}	0.0106007067137809	1	0.0106007067137809	2.97894736842105	3
10	{FwdIATMin=FwdIATMin3.0}	{IdleMin=IdleMin0}	0.0106007067137809	1	0.0106007067137809	1.04428044280443	3

Showing 1 to 10 of 234,056 entries Previous 1 2 3 4 5 … 23406 Next

Figure 7.15. Règles d'association restante de l'essai 3

En appliquant les valeurs choisis du Support = 0,008; Confiance 0,9 et le Lift=1, nous trouvons des relations pertinentes qui donnent une forte relation surtout entre l'item ATT et d'autres avec une confiance qui parfois dépasse 100% pour plusieurs relations. Cette confiance de plus de 100% est présente dans beaucoup des relations à l'exception de quelques-unes qui sont supérieures à 50%.

	lhs	rhs	support	confidence	coverage	lift	count
1	{FwdPkts.s=FwdPkts/s142857.14285714287}	{ï..BwdPkts.s=ï»¿BwdPkts/s142857.14285714287}	0.0106007067137809	1	0.0106007067137809	94.3333333333333	3
2	{ï..BwdPkts.s=ï»¿BwdPkts/s142857.14285714287}	{FwdPkts.s=FwdPkts/s142857.14285714287}	0.0106007067137809	1	0.0106007067137809	94.3333333333333	3
3	{FlowIATMin=FlowIATMin7.0,FwdIATMin=FwdIATMin0}	{FwdPkts.s=FwdPkts/s142857.14285714287}	0.0106007067137809	1	0.0106007067137809	94.3333333333333	3
4	{FlowIATMin=FlowIATMin7.0,FwdIATMax=FwdIATMax0}	{FwdPkts.s=FwdPkts/s142857.14285714287}	0.0106007067137809	1	0.0106007067137809	94.3333333333333	3
5	{FlowIATMin=FlowIATMin7.0,FwdIATStd=FwdIATStd0}	{FwdPkts.s=FwdPkts/s142857.14285714287}	0.0106007067137809	1	0.0106007067137809	94.3333333333333	3
6	{FlowIATMin=FlowIATMin7.0,FwdIATMean=FwdIATMean0}	{FwdPkts.s=FwdPkts/s142857.14285714287}	0.0106007067137809	1	0.0106007067137809	94.3333333333333	3
7	{FlowIATMin=FlowIATMin7.0,FwdIATTot=FwdIATTot0}	{FwdPkts.s=FwdPkts/s142857.14285714287}	0.0106007067137809	1	0.0106007067137809	94.3333333333333	3
8	{FlowIATMin=FlowIATMin7.0,FlowIATStd=FlowIATStd0.0}	{FwdPkts.s=FwdPkts/s142857.14285714287}	0.0106007067137809	1	0.0106007067137809	94.3333333333333	3
9	{FlowIATMin=FlowIATMin7.0,FwdIATMin=FwdIATMin0}	{ï..BwdPkts.s=ï»¿BwdPkts/s142857.14285714287}	0.0106007067137809	1	0.0106007067137809	94.3333333333333	3
10	{FlowIATMin=FlowIATMin7.0,FwdIATMax=FwdIATMax0}	{ï..BwdPkts.s=ï»¿BwdPkts/s142857.14285714287}	0.0106007067137809	1	0.0106007067137809	94.3333333333333	3

Showing 1 to 10 of 289 entries Previous 1 2 3 4 5 ... 29 Next

Figure 7.16. Règles d'association restante après élagage de l'essai 3

Nous terminons notre dernier essai en analysant la perte des informations avant et après l'élagage des règles non pertinentes comme le montrent les **Figures 7.17.1 ; 7.17.2 et 8.17.3**

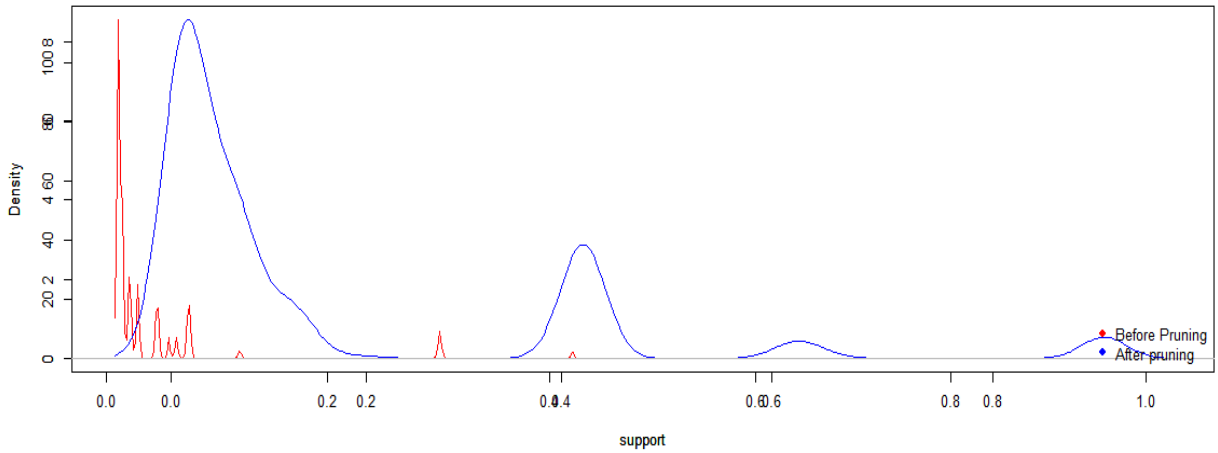

Figure 7.17.1. Densité de l'indice du support de l'essai 3

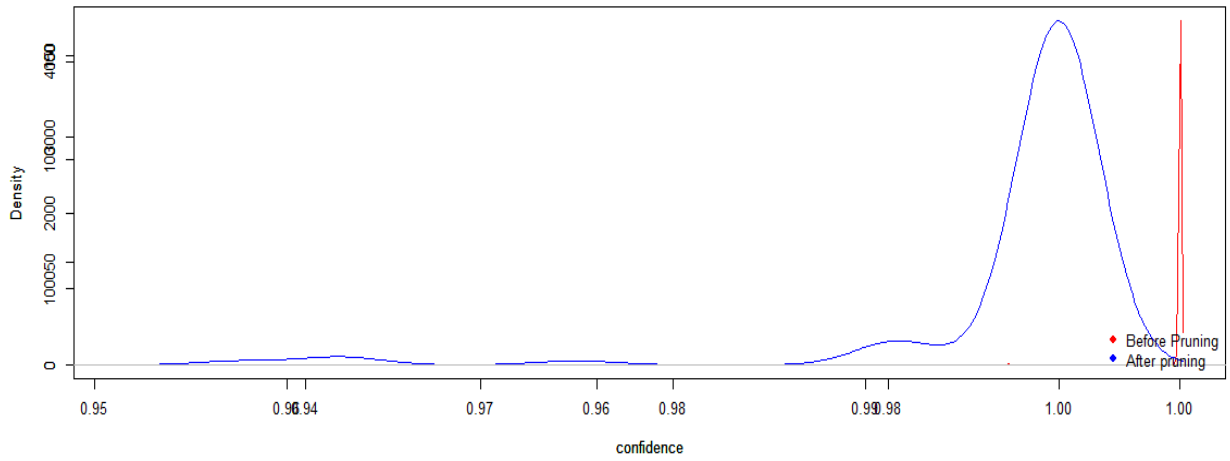

Figure 7.17.2. Densité de l'indice de la Confiance de l'essai 3

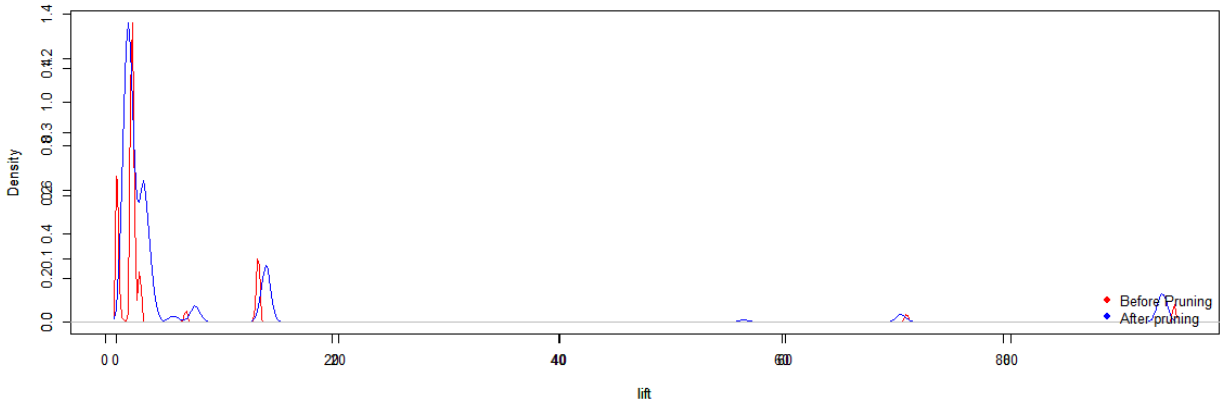

Figure 7.17.3. Densité de l'indice du Lift de l'essai 3

En faisant une lecture sur les trois (3) graphiques qui représentent chacune des métriques, nous voyons qu'il n'y a presque pas de perte importante des données entre l'avant et l'après l'élagage des règles. Avec la métrique **support**, il y a des pertes seulement au début et plus que l'on s'éloigne de 0, les pertes disparaissent tandis qu'avec la métrique **confiance**, il n'y a aucune perte entre 0,9 et 1. Enfin, avec la métrique **Lift** aussi, nous n'enregistrons presque pas de perte.

7.3.4. Discussion

Le but de nos différents essais effectués était de trouver un essai avec des indices capables de nous générer les règles d'association non seulement importante, mais aussi pertinente et qui peuvent confirmer la présence d'une attaque dans notre réseau V2G. Nous nous sommes plus attardés sur les relations qui incluent la variable **ATT**.

Nous avons constaté que le premier essai n'a pas produit de règles importantes alors que le deuxième qui a produit quelques règles importantes, mais elles n'étaient pas productrices d'informations pertinentes, car soit l'indice du **support** est bon, mais la **confiance** est inférieure au seuil choisi ou alors le **lift** n'est pas bon, ou soit c'est l'inverse. Et aussi ce deuxième essai présente au niveau du graphe a produit des pertes des informations importantes comme nous pouvons le lire aux **Figures 7.11.1, Figure 7.11.2, figure 7.11.3**.

Dans le troisième essai, nous avons obtenu des règles d'association importantes qui ont produit des informations pertinentes avec une perte des données presque négligeable et qui nous ont permis de prendre la décision finale sur les items à garder comme variables susceptible de confirmer la présence d'une attaque, d'améliorer le système d'alerte des fausses attaques.

Grâce aux résultats obtenus, nous avons réduit notre dataset qui sera utilisé dans un outil de la régression logistique afin de produire un modèle mathématique.

7.4. Régression logistique

Comme nous sommes dans la partie des essais et interprétation des résultats obtenus, nous allons alors analyser, interpréter et discuter les résultats de la régression logistique binaire obtenus dans le logiciel RStudio.

7.4.1. Chargement et exploration des données

Après avoir installé les différents packages nécessaires à l'exécution de notre programme dans RStudio, une lecture et un affichage des données de notre dataset ont été faits pour se rassurer des données que nous allons modéliser.

7.4.2. Visualisation des données

Nous avons effectué cette opération de visualisation pour résumer notre dataset à partir d'un histogramme à barre dont en hauteur est représenté le nombre d'instances qui tombe dans chacune des variables comme le montre la **Figure 7.18.**

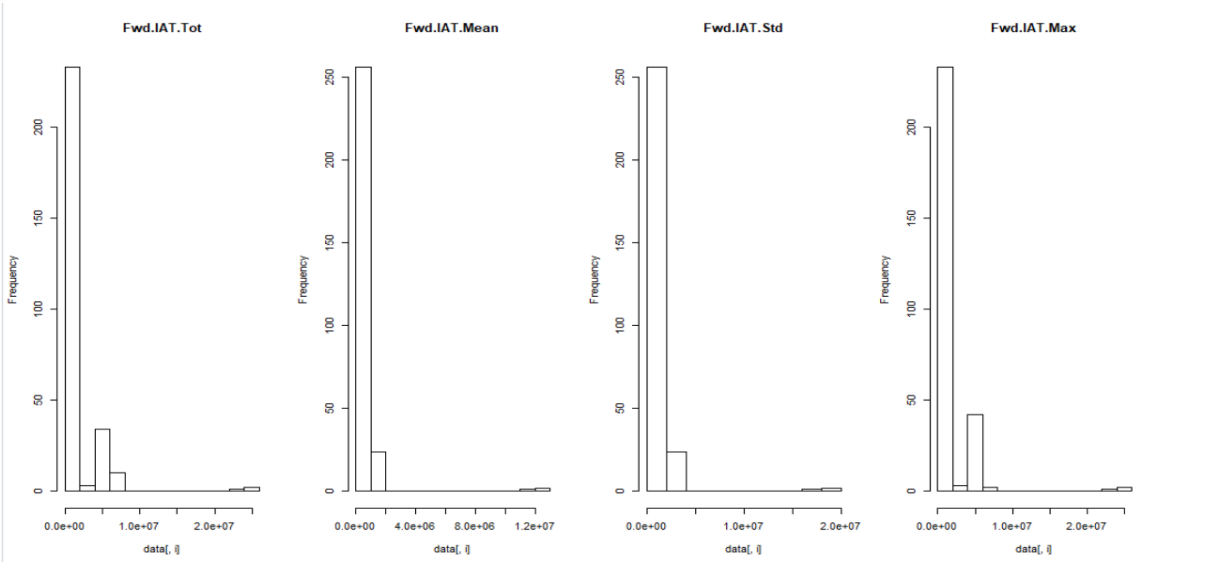

Figure 7.18. Histogramme des variables dépendantes

L'histogramme de nos variables qui monte jusqu'à la fréquence de 200 pour les variables FwdIATMax et FwdIATTot, à 250 pour FwdIATStd et FwdIATMean n'affiche pas bien les valeurs. Le niveau de variation qui varie selon les valeurs de la variable peut être complété par le graphe de distribution des données par les tracés en boite et à moustaches comme le montre la **Figure 7.19.** ci-après.

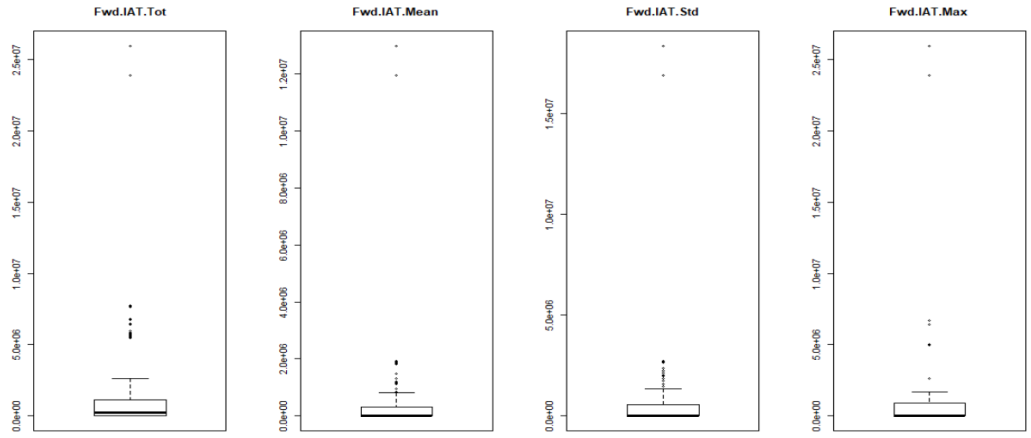

Figure 7.19. Graphe de distribution des données par les tracés en boite et à moustaches

7.4.3. Corrélation entre les variables

Dans la matrice de corrélation présentée dans la **Figure 7.20** sont affichées les corrélations entre les variables indépendantes en utilisant la couleur bleue pour marquer la présence de la corrélation (corrélation positive) et le rouge pour le manque de corrélation (corrélation négative). La grandeur des points traduit le niveau de la corrélation, car plus le point est grand, plus la corrélation est grande.

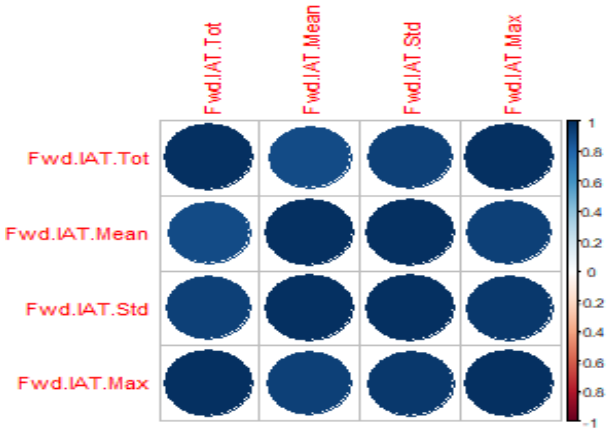

Figure 7.20. Corrélation entre les variables

Le constat est que toutes les variables sont fortement corrélées entre la variable elle-même, puis entre chacune des variables et les autres. Aucune corrélation au rouge.

En utilisant la variable dépendante **ATT** qui est binaire comme indicateur de couleur, on obtient dans le graphique des données ci-après **Figure 7.21**, une matrice de nuage de points.

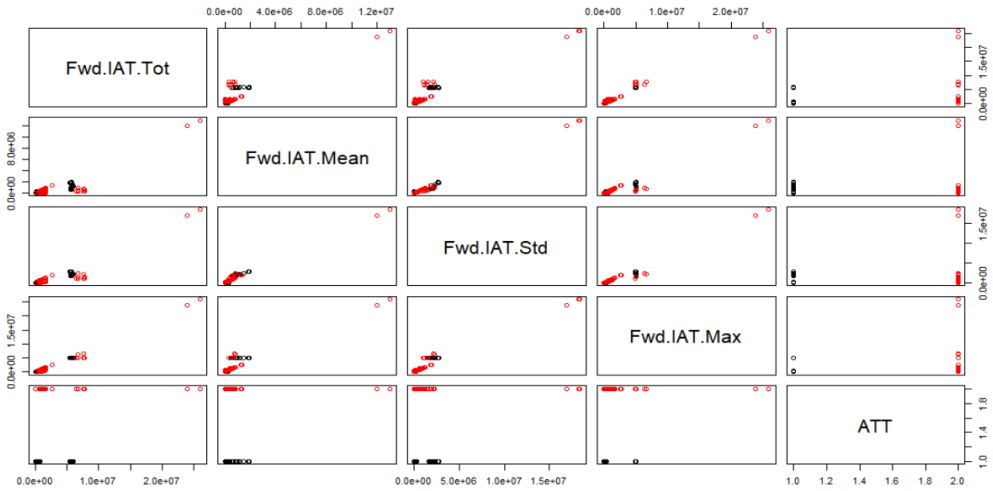

Figure 7.21. Matrice de nuage

Ce graphique montre aussi la présence des corrélations entre les variables.

En examinant la distribution de densité de chaque variable ventilée par la variable dépendante **ATT**, le diagramme de densité par **ATT** peut nous permettre de connaitre s'il y a corrélation et/ou chevauchement entre la variable ATT et les autres.

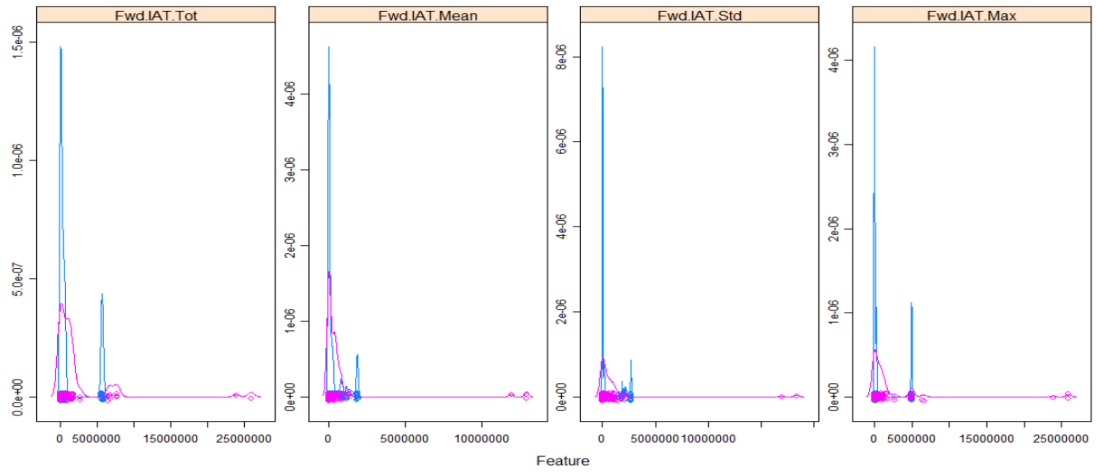

Figure 7.22. Diagramme de densité

Le graphique nous montre qu'il n'y a pas de chevauchement entre les variables, il y a au contraire une différence qui nous permet de prédire facilement s'il y a présence d'une attaque ou non.

7.4.4. Construction du modèle

La construction de notre modèle tire son contenu dans un résumé qui renvoie l'estimation, les erreurs standard, le score z, les valeurs p sur chacun des coefficients, la déviance nulle et la déviance résiduelle. Pour notre cas, 70% des données sont utilisées comme données d'entrainement et 30% comme données de test.

Nous allons aussi faire une prédiction afin de confirmer s'il y a attaque ou pas en transformant les probabilités en classifications au seuil de 0,5 qui signifie que si le vecteur est supérieur à 0,5; on confirme l'attaque ; sinon on rejette la présence de l'attaque.

Après avoir effectué cette opération, nous avons obtenu un taux de classification de 0.6904762 soit 69% qui est largement suffisant pour confirmer la positivité du modèle (Cf. **Figure 7.23**).

```
> scales <- list(x=list(relation="free"), y=list(relation="free"))
> featurePlot(x=x, y=y, plot="density", scales=scales)
> library(caret)
> index = createDataPartition(data$ATT, p = 0.70, list = FALSE)
> train = data[index, ]
> test = data[-index, ]
> # Fit a smaller model
> glm.fit = glm(ATT ~ Fwd.IAT.Max  + Fwd.IAT.Tot +  Fwd.IAT.Mean + Fwd.
IAT.Std
+                            ,data = train, family = binomial , maxit =
 100)
> glm.probs = predict(glm.fit, newdata = test, type = "response")
> glm.pred = ifelse(glm.probs > 0.5, "OUI", "NON")
> mean(glm.pred == test$ATT)
[1] 0.6904762
>
```

Figure 7.23. Prédiction du modèle

Le résumé de notre modèle est présenté à la **Figure 7.24** ci-après :

```
> summary(glm.fit)

Call:
glm(formula = ATT ~ Fwd.IAT.Max + Fwd.IAT.Tot + Fwd.IAT.Mean +
    Fwd.IAT.Std, family = binomial, data = train, maxit = 100)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2220  -0.8321  -0.8321   1.3701   1.9301

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -8.827e-01  1.921e-01  -4.594 4.34e-06 ***
Fwd.IAT.Max  -2.972e-06  1.027e-06  -2.893  0.00381 **
Fwd.IAT.Tot   1.518e-06  5.448e-07   2.787  0.00532 **
Fwd.IAT.Mean -3.601e-06  2.356e-06  -1.528  0.12645
Fwd.IAT.Std   4.885e-06  2.428e-06   2.012  0.04422 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 254.25  on 198  degrees of freedom
Residual deviance: 239.10  on 194  degrees of freedom
AIC: 249.1

Number of Fisher Scoring iterations: 5
```

Figure 7.24. Résumé du modèle

7.4.5. Discussion

Au terme de la construction de notre modèle basé sur l’optimisation du système de détection des intrusions dans le réseau V2G, la régression logistique qui a été utilisée comme méthode nous a montré qu’il y a bien une forte corrélation entre les variables de l’équation de notre modèle.

En effet, de la visualisation des données à la construction du modèle en passant par l’histogramme, la corrélation entre les variables, la matrice de nuage et le diagramme de densité, nous avons constaté la présence de fortes corrélations entre les variables et il n’y a pas de chevauchement, et donc le modèle nous donne facilement la possibilité de prédire une attaque.

Avec 70% des données utilisées pour l’entrainement et 30% données pour le test, pour un seuil de classification de 0,5; nous avons obtenu un taux de prédiction de 0,69 qui dépasse largement le seuil avec un taux de prédiction en pourcentage de 69%, cela prouve l’efficacité du modèle. Toutes les variables utilisées dans la régression logistique figurent aussi dans l’équation du modèle comme on peut le voir à la **Figure 7.23**.

En somme, nous avons analysé dans ce chapitre les résultats des différents essais de la combinaison entre les règles d'associations et la régression logistique binaire afin de montrer à quel niveau notre modèle est efficace pour l'estimation de la probabilité qu'une attaque soit vraie et d'éviter les fausses alertes.

Les données de notre dataset ont été obtenues par simulation avec l'outil MiniV2G dont les étapes et les scénarios ont été expliqués dans ce chapitre.

Les résultats obtenus dans les trois (3) essais des règles d'association nous ont permis de retenir ceux du troisième essai qui ont donné des règles importantes avec des données pertinentes occasionnant l'élagage des règles non pertinentes. Seules les règles d'association en relation avec le l'item **ATT** ont été considérées. La conséquence de cette opération nous a amené à la réduction de notre dataset qui est resté avec uniquement les items (variables) intéressants. Ensuite, ces variables ont été utilisées dans la modélisation du système réalisé avec la régression logistique exécutée dans le logiciel R.

Conclusion

Au terme de notre chapitre qui nous a permis de présenter les simulations du dataset, de réaliser trois (3) essais pour les règles d'association et d'expérimenter la régression logistique, nous pouvons retenir que les résultats obtenus sont fiables.

CHAPITRE 8 : CONCLUSION GENERALE ET PERSPECTIVES

Dans le cadre de ce mémoire de la maitrise en mathématiques et informatique, nous proposons un nouveau modèle qui améliore les systèmes de détection des intrusions dans les réseaux véhiculaires V2G en se basant sur les règles d'association maximale et la régression logistique

Nous avons pu démontrer la pertinence de l'utilisation des règles d'association maximale et la régression logistique dans un processus permettant la détection des intrusions. Pour y arriver, nous avons simulé une base de données que nous appelons dataset et qui a fait l'objet d'un traitement commençant par une analyse de corrélation de Pearson avec le logiciel SPSS avant d'appliquer le processus d'extraction des règles d'association maximale en élaguant celles qui ne sont pas pertinentes, en conservant celles qui sont intéressantes, et en visualisant la perte des informations pour aboutir à une classification. Ensuite, nous avons proposé un modèle avec la méthode statistique appelée régression logistique.

En premier lieu, nous avons fait une description des matières importantes amenant à la compréhension de ce projet de recherche. C'est les cas du réseau véhiculaire V2G, de la norme ISO15118, du système de détection d'intrusion, des règles d'association, des règles d'association maximale et de la régression logistique.

Nous avons ensuite présenté la procédure de simulation de notre dataset qui a été rendu possible grâce à l'outil MiniV2G qui inclut d'autres outils comme Mininet, RiseV2G, Wireshark et le CICFlowMeter. Tous ces outils ont été décrits ainsi que les différents scénarios utilisés avant de passer à la proposition d'un modèle efficace capable d'optimiser le système de détection d'intrusion dans le réseau V2G avec un taux de prédiction important allant jusqu'à 69%. Dans la démarche, nous avons aussi utilisé la corrélation de Pearson qui nous a permis de réduire notre dataset en ne gardant que les variables qui sont compatibles avec la variable TYPE.

Enfin, nous avons procédé à une série d'essais qui nous a permis de valider les valeurs des métriques qui devraient valider notre modèle.

Le modèle que nous avons proposé et expérimenté avec notre dataset peut être utilisé avec d'autres bases de données, à la seule condition, de respecter la procédure et les formats des données à utiliser pour éviter d'obtenir des résultats biaisés.

L'utilisation de notre modèle a donné de bons résultats pour le système de détection d'intrusions dans la reconnaissance des vraies et des fausses attaques (69%).

Nous proposons à d'autres chercheurs qui s'intéresseraient à l'amélioration des IDS de ne pas se limiter à notre solution. Une amélioration du taux de prédiction supérieur au notre aiderait mieux les scientifiques. Une autre perspective serait de proposer une solution en fonction de la combinaison des trois algorithmes afin de rendre la solution plus fiable. Un autre aspect pour améliorer notre approche est d'ajouter à notre solution la possibilité de détecter l'origine de l'attaque et l'identification de l'attaquant.

Nous proposons a d'autres chercheurs de proposer des solutions qui ont plus de réactivité lorsqu'une attaque surgit en associant des IPS (Intrusion Prevention System) afin d'obtenir des résultats capables de réagir en temps réel lorsque qu'une anomalie est détectée ou qu'une intrusion est avérée.

Références et bibliographie

1. **Institut canadien de la cybersécurité** (//www.unb.ca/cic/), Université du Nouveau-Brunswick à Fredericton, Consulté le 18 mai 2021

2. **Attanasio Luca, Conti Mauro, Donadel Denis & Turrin Federico**, *MiniV2G: An Electric Vehicle Charging Emulator*, Department of Mathematics, University of Padua, Padua, Italy, January 2021.

3. **Yousef Farhaoui**, *Evaluation des systèmes de détections et de prévention des intrusions et conception d'une nouvelle architecture des IDS*, Project The design of an online learning system in Morocco (especially rural areas), June 2015, Université Moulay Ismail, Faculty of sciences and Technics, Morocco

4. **PENGWENDE ABDOULAYE OUEDRAOGO**, *Méthode d'élagage des règles d'association et estimation de la perte d'information dans les données médicales*, mémoire a la maîtrise en mathématiques et informatique appliquées, Université du Québec à Trois-Rivières, DMI, juin 2021

5. **Dinh Thai Hoang, Ping Wang, Dusit Niyato, and Ekram Hossain**, *Charging and Discharging of Plug-In Electric Vehicles (PEVs) in Vehicle-to-Grid (V2G) Systems*: A Cyber Insurance-Based Model, cs.GT, 4 Jan 2017, Cyber insurance, plug-in electric vehicle, vehicle charging, vehicle-to-grid, Markov decision process.

6. **INTERNATIONAL STANDARD ISO 15118-1**, First edition, 2013-04-15, Reference number ISO 15118-1:2013(E), Corrected version 2013-10-01, Road Vehicles-Vehicle to grid communication interface, Part 1: General information and use-case definition

7. **Guillaume LE GALL, Nassim RIZOUG, Benjamin Cama, Sébastien SAUDRAIS et Jean-Marie Bonnin**, *Une plate-forme de recherche pour la charge intelligente de véhicules électriques*, Institut Mines-Télécom ESTACA, 11/12/2015, Les Rencontres de la Mobilité Intelligente 2016

8. **Recommandations sur le déploiement de la norme ISO 15118 en France et des services associés**, AFIREV (Association Française pour l'Itinérance de la Recharge Electrique des

Véhicules), 6 janvier 2020, in https://www.afirev.fr/wp-content/uploads/2020/01/AFIREV-Livre-blanc-d%C3%A9ploiement-15118_Vf-3.pdf, Consulté le 2 mars 2021

9. **INTERNATIONAL STANDARD ISO 15118-2**, First edition, 2014-04-01, Road vehicles - Vehicle-to-Grid Communication Interface, Part 2: Network and application protocol requirements

10. **Chakir, El Mostapha & Moughit, Mohamed & Idrissi Khamlichi, Youness,** *Building an Efficient Alert Management Model for Intrusion Detection Systems*, Advances in Science, Technology and Engineering Systems Journal (ASTESJ), 2018, 3. 18-24. 10.25046/aj030103.

11. **Saltzer, J. H., D. P. Reed, and D. D. Clark**, *End-to-End Arguments in System Design*, Proceedings of the Second International Conference on Distributed Computing Systems, Paris, France, du 8 au 10 avril 1981. IEEE Computer Society, pp. 509-512.

12. **Daniela Brauckhoff, Xenofontas Dimitropoulos, Arno Wagner, and Kave Salamatian**, *Anomaly Extraction in Backbone Networks using Association Rules*, TIK-Report 309, ftp://ftp.tik.ee.ethz.ch/pub/publications/TIK-Report-309.pdf, Consulté le 20 Septembre 2020

13. **Philippe Biondi**, Architecture expérimentale pour la détection d'intrusions dans un système informatique, Avril-Septembre 2001

14. **DOMINGUEZ Hugo et Tran Van Tay**, *Le système de détection des intrusions et le système d'empêchement des intrusions (zero day)*, Montréal, Février 2005

15. **Hatem Bouzayani**, *Modèle quantitatif pour la détection d'intrusion. Une architecture collaborative IDS-HONEYPOT*, Université de Québec en Outaouais (UQO), Mémoire de Maîtrise

16. **MESSOUAF Sonia**, *Génération automatique des scénarios d'attaques dans les systèmes informatiques*, mémoire de Master en informatique, Université de Béjaia, 2013, Algerie, Consulté le 20 Septembre 2020

17. **Mohammed S. Gadelrab, Anas Abou El Kalam, and Yves Deswarte**, *Execution Patterns in Automatic Malware and Human-centric Attacks*, NCA 2008: Proceedings of the 2008 Seventh IEEE International Symposium on Network Computing and Applications vol. 29-36.

18. **Pierre-Louis Lussan**, *Les 10 types de cyberattaques les plus courants*, Magazine: Dream Teams and How to Manage Them, septembre 2019, Netwrix Blog, https://blog.netwrix.fr/2018/07/04/les-10-types-de-cyberattaques-les-plus-courants/ Consulté le 2 mars 2021

19. **Jean-Olivier Gerphagnon, Marcelo Portes de Albuquerque, & Marcio Portes de Albuquerque**, *Sécurité Informatique, Attaques Informatique*s, Centro Brasileiro de Pesquisas F´ısicas – CBPF/CNPq, Coordena¸c˜ao de Atividade T´ecnicas – CAT, Rua Dr. Xavier Sigaud 150, 22290-180 Rio de Janeiro – RJ – Brazil,

20. **Nathalie Dagorn,** *Détection et prévention d'intrusion : présentation et limites, Rapport de recherche*, Université de Nancy1, Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA), Campus Scientifique, Vandoeuvre-lès-Nancy, France, 2006.

21. **Florian Bonne**, *Pourquoi votre stratégie de cybersécurité ne peut pas reposer sur une sonde*, TAGS : Cybersécurité Industrielle, La Cybersécurité - Par Stormshield, Publié le 12 08 2019 et Modifié le 12 08 2020, in https://www.stormshield.com/fr/actus/pourquoi-votre-strategie-de-cybersecurite-ne-peut-pas-reposer-que-sur-une-sonde/ , Consulté le 2 mars 2021

22. **Vencislav Trifonov,** *Reliability problem of intrusion protected system***,** Technical University of Sofia, Proceedings of the International Conference on Information Technologies (InfoTech-2010)16-17 September 2010, Bulgaria

23. **K. Müller**, IDS - *Systèmes de Détection d'Intrusion, Partie II*, July 2003, http://www.linuxfocus.org/Francais/July2003/article294.shtml, Consulté le 20 Septembre 2020

24. **Yergeau E. et Poirier M.,** *SPSS* à l'UdeS, 2013, in http://spss.espaceweb.usherbrooke.ca, Consulté le 15 mars 2021

25. **Lallich, S. and O. Teytaud**, *Évaluation et validation de l'intérêt des règles d'association*, Revue des nouvelles Technologies de l'information, 2003.

26. **HASSANE HILALI**, *Application de la classification textuelle pour l'extraction des règles d'association maximales*, mémoire de maîtrise en mathématiques et informatique appliquées, UQTR, DMI, avril 2009

27. **Sylvie Jami, Tao-Yan Jen, Dominique Laurent, Georges Loizou, Oumar Sy**, *Extraction de règles d'association pour la prédiction de valeurs manquantes*, Revue ARIMA, Novembre 2005, Numéro spécial CARI'04, pages 103 à 124

28. **Hacène Cherfi, Yannick Toussaint**, *Interprétation des règles d'association extraites par un processus de fouille de textes*, 13ème Congrès francophone AFRIF-AFIA de Reconnaissance des Formes et d'intelligence Artificielle, RFIA, 02-2002, Angers, France, pp.975-983, inria-00099405, Submitted on 26 Sep 2006

29. **Steve DESCÔTEAUX**, *les règles d'association maximale au service de l'interprétation des résultats de la classification*, mémoire de maitrise, UQTR, DMI, 2014

30. **Nicolas Pasquier**, *Data Mining : Algorithmes d'extraction et de réduction des règles d'association dans les bases de données*, Autre [cs.OH], Université Blaise Pascal – Clermont Ferrand II, 2000. Français, Submitted on 28 Mar 2010, Archive Ouverte Pluridisciplinaire HAL.

31. **LABIAD ALI**, *Sélection des mots clés basée sur la classification et l'extraction des règles d'association*, mémoire présenté à l'Université du Québec à Trois-Rivières à la maîtrise en mathématiques et informatique appliquées, juin 2017

32. **Marie-Jeanne Vieille**, *Règles d'association comment ça marche* ? , Lovely analytics, https://www.lovelyanalytics.com/2018/12/11/regles-dassociation-comment-ca-marche/#:~:text=Pour%20calculer%20les%20r%C3%A8gles%20de,P(F1%20%E2%88%A9%20F2). Consulté le 15 mars 2021

33. **Ricco RAKOTOMALALA**, *les règles d'association : L'analyse du panier de la ménagère*, Tutoriels Tanagra -http://tutoriels-data-mining.blogspot.fr/ , Consulté le 22 Septembre 2021

34. **Hyeok Kong, Cholyong Jong and Unhyok Ryang**, *Rare Association Rule Mining for Network Intrusion Detection*, October 2016

35. **Amir, Amihood & Aumann, Yonatan & Feldman, Ronen & Fresko, Moshe**, *Maximal Association Rules: A Tool for Mining Associations in Text*, J. Intell. Inf. Syst.. 25. 333-345. 10.1007/s10844-005-0196-9, 2005.

36. **ACHOURI ABDELGHANI**, *Extraction de relations d'associations maximales dans les textes : représentation graphique*, mémoire de maîtrise en mathématiques et informatique appliquées, université du Québec à Trois-Rivières, octobre 2012

37. **ABDERRAOUF NOUASRIA**, *Extraction d'associations lexicales fortes dans les commentaires*, mémoire de maîtrise en mathématiques et informatique appliquées, UQTR, DMI, JUIN 2016

38. **De Baripedia,** *Les régression logistiques, dernière modification le 5 mai*, Creative Commons Attribution-ShareAlike, in https://baripedia.org/wiki/Les_r%C3%A9gression_logistiques, Consulté le 15 mars 2021

39. **Avinash Navlani,** *Understanding Logistic Regression in Python Learn about Logistic regression, its basic properties, and build a machine learning model on a real-world application in Python*, December 16th, 2019

40. **François Goffinet**, *Concepts IDS IPS*, in Cisco CCNA 200-301 Volume 1, Cisco CCNA 200-301 Volume 2, Cisco CCNA 200-301 Volume 3, and Cisco CCNA 200-301 Volume 4, mise à jour le 10 août 2020,

41. **Sonia NEJI et Anne-Hélène JIGOREL**, *La régression logistique*, in chrome-extension://ohfgljdgelakfkefopgklcohadegdpjf/https://perso.univ-rennes1.fr/valerie.monbet/ExposesM2/2013/La%20re%CC%81gression%20logistique.pdf , , Consulté le 5 juillet 2020

42. **El Sanharawi, Mohamed & Naudet, Florian,** *Comprendre la régression logistique [Understanding logistic regression]*. Journal française d'ophtalmologie. 2013 Oct;36(8):710-5. French. doi: 10.1016/j.jfo.2013.05.008. Epub 2013 Aug 14. PMID: 23953846.

43. **Laroussi Karim, Amar Bensaber Boucif, Mesfioui Mhamed, Biskri Ismail**, *A probabilistic model to corroborate three attacks in Vehicular Ad Hoc Networks*, 2015, 7th IEEE International Workshop on Performance Evaluation of Communications in Distributed Systems and Web based Service Architectures

44. V2G Clarity, The Basics of Plug & Charge, ISO 15118's feature for a more user-convenient and secure way of charging electric vehicles, February 25, 2019, in https://v2g-clarity.com/knowledgebase/basics-of-plug-and-charge/, Consulté le 22 mai 2021

45. **H. Delacour1, A. Servonnet A. Perrot J.F. Vigezzi J.M. Ramirez**, *La courbe ROC (receiver operating characteristic) : principes et principales applications en biologie clinique*, revue générale, Article reçu le 9 août 2004, accepté le 30 novembre 2004, Ann Biol Clin, vol. 63, n° 2, mars-avril 2005

46. **Petros Toupas, Dimitra Chamou, Konstantinos M. Giannoutakis, Anastasios Drosou, Dimitrios Tzovaras**, *An Intrusion Detection System for Multi-Class Classification based on Deep Neural Networks,* Information Technologies Institute Center for Research & Technology-Hellas, 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA) 57001, Thessaloniki, Grece

47. **Subba, Basant & Biswas, Santosh & Karmakar, Sushanta**. *Enhancing performance of anomaly-based intrusion detection systems through dimensionality reduction using principal component analysis,* 1-6. 10.1109/ANTS.2016.7947776.

48. Université du Nouveau-Brunswick, Institut canadien pour la cybersécurité, in https://www.unb.ca/cic/datasets/nsl.html, Consulté le 22 décembre 2021

49. **Ping-ping, M., Qiu-ping**, **Z**. *Règles d'association appliquées à la détection d'intrusion.* Université de Wuhan. J. Nat. Sci. **7,** 426–430 (2002). https://doi.org/10.1007/BF02828242, Consulté le 4 Janvier 2022

50. **Subba, Basant & Biswas, Santosh & Karmakar, Sushanta**. *Intrusion Detection Systems using Linear Discriminant Analysis and Logistic Regression,* Conference Paper, Assam, India, December 2015, 1-6. 10.1109/INDICON.2015.7443533, 2015

51. **Ozge Yucel Kasap, Nevzat Ekmekci et Utku Gorkem Ketenci**, *Combining Logistic Regression Analysis and Association Rule Mining via MLR Algorithm*, Cybersoft R&D Center, Istanbul, Turkey, ICSEA 2016, The Eleventh International Conference on Software Engineering Advances

52. **Pullagura Indira Priyadarsini, G. Anuradha,** *A novel ensemble modeling for intrusion detection system,* International Journal of Electrical and Computer Engineering (IJECE), Vol. 10, No. 2, April 2020, pp. 1963~1971 ISSN: 2088-8708, DOI: 10.11591/ijece.v10i2.pp1963-1971, Accepted Nov 7, 2019

53. **SAIDOU DIOP**, *une infrastructure à clés publiques (PKI) pour sécuriser les messages dans un réseau V2G*, mémoire a la maîtrise en mathématiques et informatique appliquées, UQTR, DMI, mars 2018

54. **Pierre Paillé**, Ph.D., *La méthodologie de recherche dans un contexte de recherche professionnalisante : douze devis méthodologiques exemplaires, recherches qualitatives*, (2007), Vol. 27(2), pp. 133-151., avancées en méthodologies qualitatives, université de Sherbrooke

55. **Primavera De Filippi, Melanie Dulong de Rosnay**. *Le pirate informatique, un explorateur des courants juridiques du réseau*. Traces, École normale supérieure -lettres et sciences humaines, (ENS-LSH), 2014, pp.42. <hal-01026109>

56. **Kaur, Karamjeet & Singh, Japinder & Ghumman, Navtej**, *Mininet as Software Defined Networking Testing Platform*, Conference Paper, August 2014,

57. **F. Keti and S. Askar**, *Emulation of Software Defined Networks Using Mininet in Different Simulation Environments*, 2015 6th International Conference on Intelligent Systems, Modelling and Simulation, 2015, pp. 205-210, doi: 10.1109/ISMS.2015.46.

58. **R. L. S. de Oliveira, C. M. Schweitzer, A. A. Shinoda and Ligia Rodrigues Prete**, *Using Mininet for emulation and prototyping Software-Defined Networks*, 2014 IEEE Colombian Conference on Communications and Computing (COLCOM), 2014, pp. 1-6, doi: 10.1109/ColComCon.2014.6860404.

59. **Richard Sharpe**, **Ed Warnicke, Ulf Lamping,** *Guide de l'utilisateur de Wireshark*, Version 3.5.1 in https://www.wireshark.org/docs/wsug_html_chunked/, Consulté le 1 octobre 2020

60. **Habibi Lashkari, Arash**. *CICFlowmeter-V4.0 (formerly known as ISCXFlowMeter) is a network traffic Bi-flow generator and analyser for anomaly detection*. (2018).

https://github.com/ISCX/CICFlowMeter. 10.13140/RG.2.2.13827.20003. Consulté le 22 décembre 2021

61. https://www.unb.ca/cic/research/applications.html, Consulté le 22 décembre 2021

62. **Yiyan Jiang**, *Using Logistic Regression Model to Predict the Success of Bank Telemarketing*. International Journal on Data Science and Technology. Vol. 4, No. 1, 2018, pp. 35-41. doi: 10.11648/j.ijdst.20180401.15, Received: April 19, 2018; Accepted: May 17, 2018; Published: June 21, 2018

63. **Zhenwei Yu, Jeffrey J. P. Tsai, Fellow, IEEE, and Thomas Weigert**, *An Automatically Tuning Intrusion Detection System, IEEE transactions on systems*, MAN, and CYBERNETICS—PART B: CYBERNETICS, VOL. 37, NO. 2, APRIL 2007

64. **Ricco Rakotomalala**, *Analyse de corrélation Étude des dépendances - Variables quantitatives,* Université Lumière Lyon 2, Version 1.0, 11-Jan-2012

65. **Lenca, Philippe & Meyer, Patrick & Picouet, Philippe & Vaillant, Benoît & Lallich, Stéphane**, *Critères d'évaluation des mesures de qualité des règles d'association*, janvier 2003, mise en ligne le 23 juillet 2015, in https://www.researchgate.net/publication/278812196_Criteres_d'evaluation_des_mesures_de_qualite_des_regles_d'association, Consulté le 22 décembre 2021

66. **Daniel Rajaonasy Feno**, *Mesures de qualité des règles d'association : normalisation et caractérisation des bases, Mathématiques [math]*. Université de la Réunion, 2007. Français. fftel-00462506f

67. **HASSANE HILALI**, *Application de la classification textuelle pour l'extraction des règles d'association maximales*, mémoire a la maîtrise en mathématiques et informatique appliquée, Université du Québec, avril 2009

68. **M. El Sanharawi, F. Naudet**, *Comprendre la régression logistique*, Revue générale Journal Français d'Ophtalmologie, Volume 36, Issue 8, October 2013, Pages 710-715, in https://doi.org/10.1016/j.jfo.2013.05.008, , Consulté le 18 mai 2021

69. **X. Tan, Y. Wu, and D. H. K. Tsang,** *Pareto optimal operation of distributed battery energy storage systems for energy arbitrage under dynamic pricing*, IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 7, pp. 2103–2115, July 2016.

70. **Shweta Mishra, Himanshu Sirohia, Bhaskar Sharma and Koushik Chakraborty**, *Power generation Using Vehicle to Grid, International Journal of Pure and Applied Mathematics*, Volume 119 No. 15 2018, 1539-1550 ISSN: 1314-3395 (on-line version), url: http://www.acadpubl.eu/hub/Special Issue, Consulté le 10 fevrier 2020

71. **K. Müller**, *IDS - Systèmes de Détection d'Intrusion, Partie I*, May 2003, http://www.linuxfocus.org/Francais/May2003/article292.shtml., Consulté le 20 Septembre 2020

72. **Pierre Chifflier and Arnaud Fontaine**, *Architecture système sécurisée de sonde IDS réseau*, Agence Nationale de la Sécurité des Systèmes d'Information, 25 novembre 2014, {pierre.chifflier,arnaud.fontaine}@ssi.gouv.

73. **Julie Desjardins**, *L'analyse de régression logistique, tutorial in Quantitative Methods for Psychology*, 2007, Vol. 1(1), p. 35-41, Montreal of University

74. **Yuming Zhou and Hareton Leung**, *Empirical analysis of Object-Oriented design metrics for predicting high and low severity faults*, Member, Transactions IEEE sur l'ingénierie logicielle, vol. 32,10 octobre 2006

75. **Zuo Xiang, Patrick Seeling**, *Chapter 11- Mininet: an instant virtual network on your computer, Computing in Communication Networks From theory to practice*, Academic Press, pages 219-230, 2020, **ISBN** 978-0-12-820488-7

76. **Sébastien Dudek, Jean-Christophe Delaunay and Vincent Fargues**, *V2G Injector: Whispering to cars and charging units through the Power-Line*, Synacktiv, in https://www.sstic.org/media/SSTIC2019/SSTIC-actes/v2g_injector_playing_with_electric_cars_and_chargi/SSTIC2019-Article-v2g_injector_playing_with_electric_cars_and_charging_stations_via_powerline-dudek.pdf, Consulté le 6 janvier 2022

77. **Zonggen Yia , Don Scoffielda , Andrew Meintzb , Myungsoo Junb , Manish Mohanpurkara , Anudeep Medama**, *A Highly Efficient Control Framework for Centralized*

Residential Charging Coordination of Large Electric Vehicle Populations, Idaho National Laboratory, USA, INL/EXT-19-55352, in

78. **Ezhil Reena Joie, Kannan Thirugnanam, Mukesh Singh, Praveen Kumar**, *Distributed Active and Reactive Power Transfer for Voltage Regulation using V2G System*, 2015 4th International Conference on Electric Power and Energy Conversion Systems (EPECS), 24-26 nov. 2015, INSPEC: 15695545, 10.1109/EPECS.2015.7368522, Sharjah, Émirats arabes unis

79. **Do, Thanh-Nghi & Poulet, François**, *Régression logistique pour la classification d'images à grande échelle*, janvier 2016, Université de Can Tho, Can Tho, Vietnam,

80. **Jean Bouyer**, *Régression logistique - Modélisation des variables quantitatives*, Master, septembre 2019, Université Paris Saclay, France, HAL Id: cel-00794996

81. **Anisha Garg**, *Guide complet des règles d'association, Des algorithmes qui vous aident à magasiner plus rapidement et plus intelligemment, Towards Data Science*, 3 septembre 2018

82. **Garrett Hering**, *Impact of electric vehicles: the rise of electric vehicles reinforces the need for network integration, S&P Global Market Intelligence*, 23 septembre 2021

83. **Adeline GILLET, Yves BROSTAUX & Rodolphe PALM**, *Principaux modèles utilisés en régression logistique***,** Biotechnol. Agron. Soc. Environ. 2011 15(3), 425-433