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
TOFFA ZIDANE NONVIGNON

DÉVELOPPEMENT D'UN MODÈLE DE PRÉDICTION D'ATTAQUES BASÉ SUR LES COPULES POUR LE RÉSEAU V2G

JUIN 2022

Remerciements

Mes remerciements s'adressent :

— Tout d'abord à mon directeur de recherche le Professeur Boucif Amar Bensaber pour son accompagnement et soutient dans le cadre de la réalisation de cette recherche,
— Je remercie mon codirecteur Mhamed Mesfioui pour l'apport relatif aux notions mathématiques de ce travail,
— Je remercie les professeurs Ismail Biskri et François Meunier qui ont pris la peine d'évaluer ce travail,
— J'ai une pensée à l'endroit de tous mes collègues du département de mathématiques et d'informatique pour leur accompagnement,
— Enfin, je voudrais exprimer ma reconnaissance envers tous ceux qui m'ont soutenu de près ou de loin pour la réussite de mes études.

LISTE DES ABRÉVIATIONS

ACP : Analyse en Composantes Principales

ANFIS : Adaptive Neuro-Fuzzy Inference System

BM : Boltzmann Machine

CNN : Convolutional Neural Network

CSV : Comma-Separated Values

DNN : Deep Neural Network

DoS : Denial of Service

EV : Electric Vehicle

EVCC : Electric Vehicle Communication Controller

EVSE : Electric Vehicle Supply Equipment

ISO : International Organization for Standardization

KNN : k-nearest neighbors

MAV : Means Absolute Value

MitM : Man-in-the-Middle

MSE : Mean Squared Error

PCAP : packet capture

RMS : Root Mean Square

RNN : Recurrent Neural Network

SDI : Systèmes de Détection d'Intrusions

SE : Supply Equipment

SECC : Supply Equipment Communication Controller

SPN : Sum-Product Network

SVM : Support Vector Machine

UIT : Union Internationale des Télécommunications

VANET : Vehicular Ad-Hoc Network

V2G : Vehicule-to-Grid

V2X : Vehicule-to-Everything

RÉSUMÉ

Le réseau véhiculaire Vehicule-to-Grid (V2G) a été conçu pour assurer la recharge des véhicules électriques et protéger l'environnement en misant sur les sources d'énergies renouvelables. Ce réseau comme tout système de communication n'est pas à l'abri de menaces cybercriminelles.

Dans ce manuscrit, nous avons proposé un modèle de prédiction d'attaque qui repose sur les notions relatives aux copules afin d'améliorer la sécurité dans le réseau véhiculaire V2G. Pour réaliser notre modèle, nous avons d'abord généré une base de données d'attaques issue de trois scénarios (scénario sans attaque, scénario avec attaques de type MitM et scénario avec attaques de type DoS). Ensuite notre méthode de prédiction d'attaques basée sur les copules a été appliquée aux variables significatives de notre base de données.

Des résultats obtenus, il en résulte que notre modèle a un taux de prédiction des attaques de 96,43%.

ABSTRACT

The Vehicule-to-Grid (V2G) vehicle network was designed to charge electric vehicles and protect the environment by focusing on renewable energy sources. This network, like any communication system, is not immune to cybercrime threats.

In this manuscript, we have proposed an attack prediction model that relies on notions of copulas in order to improve security in the V2G vehicular network. To carry out our model, we first generated an attack database resulting from three scenarios (scenario without attack, scenario with MitM type attack and scenario with DoS type attack). Then our attack prediction method based on the copulas was applied to the significant variables of our database.

From the results obtained, it will result that our model has a very satisfactory attack prediction rate.

Table des matières

Table des figures

Chapitre 1

INTRODUCTION

Dans notre société actuelle, l'électricité joue un rôle si important qu'on ne peut s'en passer. Elle provient en majeur partie des sources telles que les combustibles fossiles, le nucléaire et les énergies renouvelables. Le transport de l'électricité de son lieu de production vers les usagers est réalisé par le réseau électrique. Pour répondre à la demande de l'électricité sans cesse croissante et dans le but d'optimiser l'utilisation du réseau électrique, les chercheurs ont choisi d'intégrer les systèmes d'information au réseau électrique pour obtenir un tout nouveau réseau appelé Smart Grid ou réseau intelligent qui apporte plusieurs avantages. Au nombre des avantages nous pouvons citer : la promotion des énergies renouvelables, la gestion de l'équilibre entre la production et la consommation, la réduction des coûts de consommation, la création de nouvelles opportunités de business. Mais comme tout réseau informatique, le Smart Grid et plus précisément le réseau Vehicule-to-Grid est confronté à des risques de piratage informatique. Le réseau Vehicule-to-Grid est une technologie ou un composant du Smart Grid qui se base sur une communication bidirectionnelle entre le véhicule électrique et sa borne de recharge. Les différents travaux effectués sur les réseaux V2G sont assez récents tout comme pour le réseau Smart Grid et pour relever les nombreux défis sécuritaires d'autres recherches doivent être effectuées. Il faudrait alors trouver des solutions innovantes afin de lutter efficacement contre les menaces potentielles au sein du réseau Vehicule-to-Grid (V2G). Pour y arriver, il est possible de s'inspirer des Systèmes de Détection d'Intrusions (SDI) basés sur les réseaux de neurones ou des modèles mathématiques. Ces Systèmes de Détection d'Intrusions ont fait leur preuve en renforçant la sécurité dans plusieurs types

de réseaux (le réseau internet, intranet, VANET, ...) suite à la recrudescence des actes de piratage informatique observés ces dernières années. Dans ce manuscrit, nous proposons un modèle de prédiction d'attaques dans le réseau V2G basé sur des notions mathématiques relatives aux copules. La phase de recherche débute par les différentes simulations du réseau V2G qui ont permis d'obtenir une base de données d'attaques. Une fois notre base de données obtenue, des opérations de nettoyage ont été effectuées sur ladite base afin de la rendre utilisable par notre modèle. Grâce au logiciel R, nous avons implémenté notre modèle basé sur les copules. Dans les chapitres 2 et 3 du présent mémoire, nous présentons respectivement les notions de copules et quelques généralités sur les réseaux V2G. Ensuite dans le chapitre 4, nous exposons les travaux de quelques chercheurs sur la sécurité des réseaux et des différentes techniques pour la détection d'attaques. Le chapitre 5 lève le voile sur la méthodologie utilisée dans le cadre de la réalisation de notre modèle. Enfin dans les chapitres 6 et 7, nous présentons respectivement notre article scientifique et les résultats obtenus.

Chapitre 2

LES COPULES

En statistique, le concept des copules joue un rôle important pour modéliser des données multivariées. Dans ce chapitre, nous allons faire une synthèse sur les copules. Ces dernières vont être utilisées pour construire notre modèle de prédiction.

2.1 Copule bivariée

Dans le présent sous-chapitre, nous exposerons les propriétés d'une copule bivariée.

2.1.1 Défintion, différentiabilité et continuité d'une copule bivariée

Définition 1.1

Soit I l'intervalle [0, 1]. On appelle copule C la fonction de $I^2 \longrightarrow I$ définie par les caractéristiques ci-après [2] :

1. $C(u,0) = C(0,u) = 0, \forall\, u \in I$;
2. $C(u,1) = C(1,u) = u, \forall\, u \in I$;
3. C *est* $2-$*croissante* : *pour tout* $u_1, u_2, v_1, v_2 \in I$ avec $0 \leq u_1 \leq v_1 \leq 1$ et $0 \leq u_2 \leq v_2 \leq 1$ on a :
   $C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0.$

Théorème 1.2 (Différentiabilité) [3]

La dérivée partielle de C par rapport à u notée $\frac{\partial C}{\partial u}$ satisfait :

$$0 \leq \frac{\partial C(u,v)}{\partial u} \leq 1.$$

Aussi $\forall\, u \in I$, $\frac{\partial C}{\partial v}$ existe pour tout $v \in I$, et on a :

$$0 \leq \frac{\partial C(u,v)}{\partial v} \leq 1.$$

Notons que sur l'intervalle $I = [0,1]$, les fonctions $u \longmapsto \frac{\partial C(u,v)}{\partial v}$ et $v \longmapsto \frac{\partial C(u,v)}{\partial u}$ sont définies et croissantes presque partout.

Théorème 1.3 (Continuité) [4]

Considérons la copule bivariée C. $\forall\, u_1, u_2, v_1, v_2 \in I$ tel que $u_1 \leq u_2$ et $v_1 \leq v_2$ nous avons :

$$|C(u_1, v_1) - C(u_2, v_2)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

2.1.2 Théorème de Sklar

Le théorème de Sklar permet de modéliser des phénomènes qui impliquent un couple de variables aléatoires (X, Y).

Proposition 1.4 [1]

1. F^{-1} représente la fonction inverse de F. La fonction, $F^{-1}(t) = \inf\{x : F(x) \geq t\} = \sup\{x : F(x) \leq t\}$ est croissante et continue à gauche. Aussi, $\forall x \in \mathbb{R}, p \in]0,1]$ $F(x) \geq p \iff x \geq F^{-1}(p)$.
2. Soit une variable aléatoire X de fonction de répartition F. Si F est continue, alors $F(X)$ suit une loi uniforme sur I.

Théorème 1.5(Sklar) [5]

Soit F une fonction de répartition bivariée dont les fonctions de répartitions marginales sont F_1 et F_2.

Et posons $F_1(x_1) = u$ *et* $F_2(x_2) = v$. L'expression de $F(x_1, x_2)$ peut s'écrire en fonction de la copule $C(u, v)$ comme suit :

$$\forall (x_1, x_2) \in \mathbb{R}^2 \,;\, F(x_1, x_2) = C(F_1(x_1), F_2(x_2)).$$

Si les marges F_1 et F_2 sont continues on conclut que la copule C est unique.

De manière réciproque, si une copule C et F_1, F_2 sont des distributions univariées, alors F définie par $C(F_1(x_1), F_2(x_2))$ est la distribution conjointe ayant pour marginales F_1 et F_2.

Preuve [5]

En partant du prinicipe que F_1 et F_2 sont continues en utilisant la condition **2.** de la proposition 1.4, $F_1(X_1)$ et $F_2(X_2)$ sont uniformes sur l'intervalle I.

Soit F la fonction de répartition du couple $F_1(X_1)$ et $F_2(X_2)$, c'est-à-dire,

$$\begin{aligned} C(u, v) &= P(F_1(X_1) \leq u, F_2(X_2) \leq v) \\ &= P(X_1 \leq F_1^{-1}(u), X_2 \leq F_2^{-1}(v)) \\ &= F(F_1^{-1}(u), F_2^{-1}(v)). \end{aligned}$$

Cela termine la démonstration de 1.4.

En se basant sur la condition **1.** de la proposition 1.4 on a :

$$\begin{aligned} F(x_1, x_2) &= P(X_1 \leq x_1, X_2 \leq x_2) \\ &= P(F_1(X_1) \leq F_1(x_1), F_2(X_2) \leq F_2(x_2)) \\ &= C(F_1(x_1), F_2(x_2)). \end{aligned}$$

Cela achève la démonstration de la réciproque du théorème.

Corollaire 1.6 (inversion de Sklar) [6]

Soient C, F, F_1 et F_2 du théorème 1.5. Supposons que F_1 et F_2 sont continues si

F_1^{-1} et F_2^{-1} représentent respectivement les fonctions inverses de F_1 et F_2. $\forall (u,v) \in I^2$, nous avons :

$$C(u,v) = F(F_1^{-1}(u), F_2^{-1}(v)).$$

Exemple [1]

Soit la distribution logistique bivariée de Gumbel [7]. Sa fonction de répartition conjointe est définie par :

$$F(x_1,x_2) = (1+e^{-x_1}+e^{-x_2})^{-1}, \text{ pour tout } (x_1,x_2) \in \mathbb{R}^2.$$

Ainsi, en calculant $\lim_{x_2 \to \infty} F(x_1,x_2)$ et $\lim_{x_1 \to \infty} F(x_1,x_2)$, on obtient les fonctions $F_1(x_1) = (1+e^{-x_1})^{-1}$ *et* $F_2(x_2) = (1+e^{-x_2})^{-1}$ qui sont des fonctions de répartitions marginales.

Les fonctions inverses sont respectivement :

$$F_1^{-1}(u) = -\ln(\tfrac{1}{u}-1) \quad et \quad F_2^{-1}(v) = -\ln(\tfrac{1}{v}-1).$$

En appliquant le théorème 1.5, la copule C associée à $F(x_1,x_2)$ s'écrit :

$$\begin{aligned} C(u,v) &= F(F_1^{-1}(u), F_2^{-1}(v)) \\ &= [1+e^{\ln(\frac{1}{u}-1)}+e^{\ln(\frac{1}{v}-1)}]^{-1} \\ &= \frac{uv}{u+v-uv}. \end{aligned}$$

2.1.3 Bornes de Fréchet-Hoeffding

Soient les fonctions suivantes introduites dans [8] :

$$W = C^-(u,v) = \max(u+v-1,0) \text{ et } M = C^+(u,v) = \min(u,v)$$

Pour toute C et pour tout $(u,v) \in [0,1]^2$, on a :

$$C^-(u,v) \leq C(u,v) \leq C^+(u,v).$$

La borne inférieure W et la borne supérieure M représentent les bornes de Fréchet-Hoeffding. Mais, il est important de noter que C^- est une copule uniquement dans le cas bivarié.

Propriété 1.7 Pour toute copule C et pour tout $(u, v) \in [0, 1]^2$, nous avons :

$$C^-(u, v) \leq C(u, v) \leq C^+(u, v).$$

Preuve

Soit $(u, v) \in I^2$ on a :

$$\begin{cases} C(u, v) \leq C(u, 1) = u \\ \text{et} \\ C(u, v) \leq C(1, v) = v \end{cases}$$

alors,

$$C(u, v) \leq \min(u, v) = C^+(u, v)$$

D'où la deuxième inégalité.

Pour montrer la première inégalité, nous nous basons sur le fait que $C(u, v) \geq 0$ et l'inégalité $C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0$.

Ainsi, nous avons :

$$\begin{aligned} C(1,1) - C(1,v) - C(u,1) + C(u,v) \geq 0 &\implies 1 - v - u + C(u,v) \geq 0 \\ &\implies C(u,v) \geq u + v - 1 \\ &\implies C(u,v) \geq \max(0, u + v - 1) \\ &\implies C(u,v) \geq C^-(u,v). \end{aligned}$$

2.1.4 Densité d'une copule

Considérons f la densité d'une loi bivariée. (voir [2]). f peut s'écrire en fonction des densités des marginales f_1 et f_2 et de la densité c de la copule associée :

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2)) \times f_1(x_1) \times f_2(x_2),$$

où c représente la densité de la copule C définie de la manière suivante :

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v).$$

L'expression de la densité de la copule si la fonction de répartition possède une densité continue est la suivante :

$$c(u, v) = \frac{f(F_1^{-1}(u), F_2^{-1}(v))}{f_1(F_1^{-1}(u)) f_2(F_2^{-1}(v))} \ .$$

2.1.5 Relation d'ordre sur les copules

Définition 1.8 [5]

Soient deux copules C_1 et C_2. On définit l'ordre $\prec$, comme suit :

$C_1 \prec C_2$ si et seulement si :

$$C_1(u, v) \leq C_2(u, v) \qquad \text{pour tout } (u, v) \in I^2.$$

Une comparaison entre toutes les copules est impossible. Pour cette raison, la relation d'ordre ci-dessus est dite partielle. Mais, la relation ci-dessous est toujours vérifiée :

$$C^- \prec C \prec C^+.$$

Lorsqu'une copule C vérifie l'inégalité $C^\perp \prec C \prec C^+$, nous concluons que cette fonction représente une structure de dépendance positive où $C^\perp$ représente la copule indépendance.

De plus, une copule C qui satisfait l'inégalité $C^- \prec C \prec C^\perp$ est nommée une structure de dépendance négative.

Par ailleurs, cette relation d'ordre étant partielle, on peut rencontrer des fonctions copules C comme $C \not\succeq C^\perp$ et $C \not\preceq C^\perp$, c'est-à-dire, qu'il existe des fonctions copules qui ne sont ni des structures de dépendance positive, ni des structures de dépendance négative.

2.1.6 Propriété d'invariance

Dans le concept des copules, l'invariance par transformations strictement croissantes est un théorème essentiel.

Proposition 1.9 [8]

Soient deux variables aléatoires continues X_1 et X_2 dont la copule associée est C_{X_1,X_2}.

Si α et β sont deux fonctions strictement croissantes sur $\mathrm{Im}(X_1)$, $\mathrm{Im}(X_2)$ respectivement alors :

$$C_{\alpha(X_1),\beta(X_2)} = C_{X_1,X_2}.$$

Proposition 1.10 [6] [5]

Soient deux variables aléatoires continues X_1 et X_2 ayant pour copule associée C_{X_1,X_2}.

Soient α et β deux fonctions strictement monotones, respectivement sur $\mathrm{Im}(X_1)$, $\mathrm{Im}(X_2)$.

1. Si α est croissante et β décroissante alors,

$$C_{\alpha(X_1)\beta(X_2)}(u,v) = u - C_{X_1X_2}(u,1-v).$$

2. Si α est décroissante et β croissante alors,

$$C_{\alpha(X_1)\beta(X_2)}(u,v) = v - C_{X_1X_2}(1-u,v).$$

3. Si α et β sont décroissantes alors,

$$C_{\alpha(X_1)\beta(X_2)}(u,v) = u + v - 1 + C_{X_1X_2}(1-u, 1-v).$$

Preuve de la proposition 1.9 (voir[5])

Soient F et G les fonctions de répartition conjointes respectives des vecteurs aléatoires (X_1, X_2) et $(\alpha(X_1), \beta(X_2))$.

Soient F_1, F_2 les fonctions marginales de F et G_1, G_2 les fonctions marginales de G

Nous remarquons que les marges de G sont :

$$\begin{aligned} G_1(x_1) &= P(\alpha(X_1) \leq x_1) \\ &= P(X_1 \leq \alpha^{-1}(x_1)) \\ &= F_1(\alpha^{-1}(x_1)), \end{aligned}$$

car α est une fonction croissante et

$$G_2(x_2) = F_2(\beta^{-1}(x_2)).$$

Nous avons donc $G_1^{-1}(u) = \alpha(F_1^{-1}(u))$ et $G_2^{-1}(v) = \beta(F_2^{-1}(v))$.

Nous en déduisons le résultat suivant :

$$
\begin{aligned}
C_{\alpha(X_1)\beta(X_2)}(u,v) &= C(G_1^{-1}(u), G_2^{-1}(v)) \\
&= P(\alpha(X_1) \leq G_1^{-1}(u), \beta(X_2) \leq G_2^{-1}(v)) \\
&= P(X_1 \leq \alpha^{-1}(G_1^{-1}(u)), X_2 \leq \beta^{-1}(G_2^{-1}(v))) \\
&= P(X_1 \leq F_1^{-1}(u), X_2 \leq F_2^{-1}(v)) \\
&= C_{X_1X_2}(u,v).
\end{aligned}
$$

Exemple [1] Nous avons :

$$
\begin{aligned}
C_{X_1,X_2} &= C_{\ln X_1, X_2} \\
&= C_{\ln X_1, \ln X_2} \\
&= C_{X_1, \exp(X_2)} \\
&= C_{\sqrt{X_1}, \exp(X_2)}.
\end{aligned}
$$

Nous concluons que l'application de transformations croissantes ne modifie pas la copule, mais les lois marginales changent.

2.1.7 Relation de dépendance

1. Fonction de concordance

Soit un vecteur aléatoire continu (X, Y) à qui on associe les réalisations (x, y) et $(\tilde{x}, \tilde{y})$. Les deux réalisations (x, y) et $(\tilde{x}, \tilde{y})$ sont dites (voir [8]) :

— concordantes si $(x < \tilde{x}$ et $y < \tilde{y})$ ou $(x > \tilde{x}$ et $y > \tilde{y})$;

— discordantes si $(x < \tilde{x}$ et $y > \tilde{y})$ ou $(x > \tilde{x}$ et $y < \tilde{y})$.

Définition 1.11 [8]

La fonction de concordance entre les couples (X, Y) et $(\tilde{X}, \tilde{Y})$ qui sont des couples de variables aléatoires continues, est la fonction Q donnée par :

$Q = P[(X - \tilde{X}) - (Y - \tilde{Y}) > 0] - P[(X - \tilde{X}) - (Y - \tilde{Y}) < 0]$.

Autrement dit, elle représente la différence entre la probabilité de concordance et celle de discordance.

2. Les propriétés de la fonction de concordance

Théorème 1.12 [4]

Soient deux couples de variables aléatoires indépendants (X, Y) et $(\tilde{X}, \tilde{Y})$ de fonctions de répartitions conjointes F et $\tilde{F}$ avec des marges communes F_1 et F_2 respectivement. Soient C_1 et C_2 les copules associées aux fonctions de répartitions F et $\tilde{F}$ respectivement alors,

$$Q = Q(C_1, C_2) = 4 \int_0^1 \int_0^1 C_2(u, v)\, dC_1(u, v) - 1.$$

Preuve [1]

Posons $u = F_1(x)$ et $v = F_2(y)$.

$$\begin{aligned} Q &= P[(X - \tilde{X}) - (Y - \tilde{Y}) > 0] - P[(X - \tilde{X})(Y - \tilde{Y}) < 0] \\ &= P[(X - \tilde{X}) - (Y - \tilde{Y}) > 0] - (1 - P[(X - \tilde{X})(Y - \tilde{Y}) > 0]) \\ &= 2P[(X - \tilde{X}) - (Y - \tilde{Y}) > 0] - 1, \end{aligned}$$

avec

$$P[(X - \tilde{X})(Y - \tilde{Y}) > 0] = P(X < \tilde{X}, Y < \tilde{Y}) + P(X > \tilde{X}, Y > \tilde{Y}).$$

Par ailleurs,

$$\begin{aligned} P(X > \tilde{X}, Y > \tilde{Y}) &= P(\tilde{X} < X, \tilde{Y} < Y) \\ &= \iint_{\mathbb{R}^2} P[\tilde{X} < x, \tilde{Y} < y] dC_1[F_1(x), F_2(y)] \\ &= \iint_{\mathbb{R}^2} C_2[F_1(x), F_2(y)] dC_1[F_1(x), F_2(y)] \\ &= \iint_{I^2} C_2(u, v) dC_1(u, v). \end{aligned}$$

De façon similaire, nous avons :

$$\begin{aligned} P(X < \tilde{X}, Y < \tilde{Y}) &= P(\tilde{X} > X, \tilde{Y} > Y) \\ &= \iint_{\mathbb{R}^2} P[\tilde{X} > x, \tilde{Y} > y] dC_1[F_1(x), F_2(y)] \\ &= \iint_{\mathbb{R}^2} \bar{\tilde{F}}(x, y) dC_1[F_1(x), F_2(y)] \qquad \bar{\tilde{F}} \text{ fonction de survie de } \tilde{F} \\ &= \iint_{\mathbb{R}^2} [1 - F_1(x) - F_2(y) + C_2(F_1(x), F_2(y))] dC_1[F_1(x), F_2(y)] \\ &= \iint_{I^2} [1 - u - v + C_2(u, v)] dC_1(u, v), \end{aligned}$$

où C_1 représente la fonction de répartition du couple (U, V) d'une loi uniforme $(0, 1)$. Notons que $E(U) = E(V) = \frac{1}{2}$ ce qui nous donne :

$$\begin{aligned} P(X < \tilde{X}, Y < \tilde{Y}) &= 1 - \tfrac{1}{2} - \tfrac{1}{2} + \iint_{I^2} C_2(u, v) dC_1(u, v) \\ &= \iint_{I^2} C_2(u, v) dC_1(u, v). \end{aligned}$$

Par suite :

$$Q = 4 \iint_{I^2} C_2(u, v) dC_1(u, v) - 1.$$

Corollaire 1.13 [5]

Soient C_1, C_2 et Q du théorème 1.12, alors,

1. Q est symétrique signifie : $Q(C_1, C_2) = Q(C_2, C_1)$
2. Q concerve l'ordre, c'est-à-dire : Si $C_1 \prec C_1'$ *et* $C_2 \prec C_2'$ $\forall (u, v) \in I^2$ alors,
$$Q(C_1, C_2) \leq Q(C_1', C_2').$$

Pour les copules usuelles W, M et Π, il est possible d'évaluer Q. En effet,

$$Q(M, M) = 1,$$

$$Q(M, \Pi) = Q(W, \Pi) = \tfrac{1}{3},$$

$$Q(M, W) = Q(\Pi, \Pi) = 0,$$

$$Q(W, W) = -1.$$

Pour toute copule C, on a :

$$\begin{aligned} 0 &\leq Q(C, M) \leq 1, \\ -1 &\leq Q(C, W) \leq 1, \\ \tfrac{-1}{3} &\leq Q(C, \Pi) \leq \tfrac{1}{3}. \end{aligned}$$

Preuve [1]

1. Le support de la copule M est l'ensemble $D_M = \{(u, v) \in I^2 \backslash u = v\}$. Rapellons que $M(u, v) = \min(u, v), W(u, v) = \max(u + v - 1, 0)$ *et* $\Pi(u, v) = uv$

- $$\begin{aligned} Q(M,M) &= 4\iint_{I^2} M(u,v)dM(u,v) - 1 \\ &= 4\iint_{I^2} M(u,v)dM(u,v) - 1 \; car \; D_M = \{(u,v) \in I^2 \backslash u = v\} \\ &= 4\int_0^1 udv - 1 = 1. \end{aligned}$$

- $$\begin{aligned} Q(M,\Pi) &= 4\iint_{I^2} \Pi(u,v)dM(u,v) - 1 \\ &= 4\iint_{I^2} u^2 du - 1 \\ &= \tfrac{1}{3}. \end{aligned}$$

- $$\begin{aligned} Q(M,W) &= 4\iint_{I^2} W(u,v)dM(u,v) - 1 \\ &= 4\iint_{I^2} (2u-1)du - 1 \\ &= 0. \end{aligned}$$

2. De même, le support de la copule W est :$D_M = \{(u,v) \in I^2 \backslash u = 1 - v\}$

- $$\begin{aligned} Q(W,\Pi) &= 4\iint_{I^2} \Pi(u,v)dW(u,v) - 1 \\ &= 4\iint_{I^2} u(1-u)du - 1 \\ &= -\tfrac{1}{3}. \end{aligned}$$

- $$\begin{aligned} Q(W,W) &= 4\iint_{I^2} W(u,v)dW(u,v) - 1 \\ &= 4\iint_{I^2} 0du - 1 \\ &= -1. \end{aligned}$$

3. Finalement, puisque $d\Pi(u,v) = dudv$ on a :

$$\begin{aligned} Q(\Pi,\Pi) &= 4\iint_{I^2} \Pi(u,v)d\Pi(u,v) - 1 \\ &= 4\iint_{I^2} uvdudv - 1 \\ &= 0. \end{aligned}$$

4. Pour toute copule C on a :

$$W \leq C \leq M \Rightarrow Q(W,M) \leq Q(C,M) \leq Q(M,M)) \Rightarrow 0 \leq Q(C,M) \leq 1$$

De la même manière, on établit que :

$$-1 \leq Q(C,W) \leq 1 \; et \; \tfrac{-1}{3} \leq Q(C,\Pi) \leq \tfrac{1}{3}.$$

3. Mesures de concordance

Théorème 1.14 [2]

Considérons X et Y des variables aléatoires continues.

On dit que κ est une mesure de concordance pour les variables X et Y si elle satisfait les propriétés suivantes :

1. si Y est une fonction croissante de X alors, $\kappa\ (X,Y) = 1$,
2. si Y est une fonction décroissante de X alors, $\kappa\ (X,Y) = -1$,
3. si α et β sont des fonctions strictement croissantes alors, $\kappa(\alpha(X),\beta(Y)) = \kappa(X,Y)$.

4. Tau de Kendall et rho de Spearman

Afin d'étudier la dépendance entre deux variables, on fait appel à plusieurs mesures. Au nombre de ces mesures, on trouve le tau de Kendall et le rho de Spearman. Nous allons d'abord introduire le tau de Kendall ainsi que ses propriétés.

a) Tau de Kendall

Définition 1.15 [2]

Soient (X, Y) et (X', Y') deux vecteurs aléatoires indépendants et de même loi.

Le tau de Kendall relatif au vecteur (X, Y) est défini de la manière suivante :

$$\tau(X, Y) = P[(X - X')(Y - Y') > 0] - P[(X - X')(Y - Y') < 0].$$

Les propriétés du tau de Kendall sont les suivantes :

- Il est symétrique, c'est-à-dire : $\tau(X, Y) = \tau(Y, X)$;
- $-1 \leq \tau \leq +1$;
- Si X et Y sont co-monotones alors, $\tau = 1$ (la concordance parfaite) ;
- Si X et Y sont contre-monotones alors, $\tau = -1$ (la discordance parfaite) ;
- Si X et Y sont indépendantes alors, $\tau = 0$ (l'inverse pas forcément vraie) ;
- Si α et β sont des fonctions strictement croissantes, $\tau(\alpha(X), \beta(Y)) = \tau(X, Y)$.

Il est possible de déduire le tau de Kendall à partir de la copule (voir la proposition suivante).

Proposition 1.16 [2]

Soit (X, Y) un vecteur aléatoire continu avec C pour copule associée.

L'expression du tau de Kendall $\tau(X, Y)$ est la suivante :

$$\begin{aligned} \tau(X, Y) &= 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) du dv - 1 \\ &= 4E[C(U, V)] - 1. \end{aligned}$$

Preuve [2] On a :

$$\begin{aligned} U &= F_X(X) \\ U' &= F_X(X') \\ V &= F_Y(Y) \\ V' &= F_Y(Y') \end{aligned}$$

dont les lois sont uniformes sur l'intervalle I.

On a :

$$\begin{aligned}
\tau(X,Y) &= P[(X-X^{'})(Y-Y^{'})\geq 0] - P[(X-X^{'})(Y-Y^{'})<0] \\
&= 2P[(X-X^{'})(Y-Y^{'})\geq 0] - 1 \\
&= 2P[(X-X^{'})\geq 0,(Y-Y^{'})\geq 0] + 2P[(X-X^{'})<0,(Y-Y^{'})<0] - 1 \\
&= 4P[X\geq X^{'},Y\geq Y^{'}] - 1 \\
&= 4P[F_X(X)\geq F_X(X^{'}),F_Y(Y)\geq F_Y(Y^{'})] - 1 \\
&= 4P[U\geq U^{'},V\geq V^{'}] - 1 \\
&= 4\int_0^1\int_0^u\int_0^1\int_0^v c(u,v)c(u^{'},v^{'})dudvdu^{'}dv^{'} - 1 \\
&= 4\int_0^1\int_0^1 C(u,v)c(u,v)dudv - 1 \\
&= 4E(C(U,V)) - 1.
\end{aligned}$$

Un estimateur empirique du tau de Kendall peut être construit si nous avons un échantillon d'observations de taille n de (X,Y), $(x_i,y_i)_{1\leq i\leq n}$. Cet estimateur est donné par [2] :

$$\hat{\tau}_n(X,Y) = \frac{2}{n(n-1)}\sum_{j=2}^{n}\sum_{i=1}^{j-1}\text{sign}[(x_j-x_i)(y_j-y_i)],$$

où la fonction **sign** est définit par :

$$\begin{cases} 1\ si\ z>0, \\ -1\ si\ z<0, \\ 0\ si\ z=0. \end{cases}$$

b) Rho de Spearman

Définition 1.17 [2]

Le rho de Spearman relatif aux variables X et Y est donné par :

$$\rho_S(X,Y) = 3(P[(X-\tilde{X})(Y-Y') > 0] - P[(X-\tilde{X})(Y-Y') < 0]),$$

où (X,Y), $(\tilde{X},\tilde{Y})et(X',Y')$ sont des couples indépendants.

Dans le cas où X et Y sont uniformes, il est possible de déterminer le rho de Spearman en fonction du coefficient de corrélation relatif à X et Y (voir proposition suivante).

Proposition 1.18 [2]

Le rho de Spearman relatif aux variables X et Y est égal au coefficient de corrélation entre les variables $F_1(X)$ et $F_2(Y)$.

$$\rho_S(X,Y) = \rho(F_1(X), F_2(Y)).$$

Preuve [2]

D'après Schweizer et Wolff on a :

$$\begin{aligned}\rho_S(X,Y) &= 12\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[F(x,y) - F_1(x)F_2(y)]dxdy \\ &= \frac{1}{\sqrt{Var(F_1(X))Var(F_2(Y))}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[F(x,y) - F_1(x)F_2(y)]dxdy,\end{aligned}$$

ce qui démontre la proposition.

Notons que l'expression du rho de Spearman peut être en fonction de la copule (voir la proposition suivante) :

Proposition 1.19 [2]

Soit (X,Y) un couple de variable aléatoire continue de copule C, l'expressions du rho de Spearman $\rho_S(X,Y)$ est la suivante :

$$\begin{aligned}\rho_S(X,Y) &= 12\int_0^1\int_0^1 C(u,v)dudv - 3 \\ &= 12\int_0^1\int_0^1 uvdC(u,v) - 3.\end{aligned}$$

Preuve On considère deux variables $U = F_X(X)$, et $V = F_Y(Y)$ suivant une loi uniforme sur I. Par conséquent, nous avons :

$$E(U) \ = E(V) = \tfrac{1}{2},$$

et

$$Var(U) \ = Var(V) = \tfrac{1}{12}.$$

D'une part, en se basant sur la définition du coefficient de corrélation, on trouve,

$$\begin{aligned}
\rho_S(X,Y) &= \rho(F_1(X), F_2(Y)) \\
&= \frac{1}{\sqrt{Var(F_1(X))Var(F_2(Y))}} \int_0^1 \int_0^1 (C(u,v) - uv) du dv \\
&= 12 \int_0^1 \int_0^1 C(u,v) du dv - 3.
\end{aligned}$$

D'autre part on a :

$$\begin{aligned}
\rho(X,Y) &= \rho(U,V) \\
&= \frac{\mathrm{cov}(U,V)}{\sigma_U \sigma_V} \\
&= 12\mathrm{cov}(U,V) \\
&= 12 \int_0^1 \int_0^1 uv dC(u,v) - 3.
\end{aligned}$$

2.2 Copule multidimensionnelle

Dans cette section, nous allons exposer certaines propriétés importantes des copules multidimensionnelles.

2.2.1 Définition d'une copule multivariée

Définition 1.20 [5]

Une copule n-dimensionnelle est une fonction ayant les caractéristiques suivantes :

i. Pour tout vecteur aléatoire $u = (u_1, \ldots, u_n) \in I^n, C(u) = 0$ si nous avons au moins une coordonnée de u est égale à 0.

ii. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i, \forall u_i \in I, i = 1, \ldots, n.$

iii. C est n-croissante, c'est-à-dire $\forall u = (u_1, \ldots, u_n), v = (v_1, \ldots, v_n)$ dans I^n tels que $u_i \leq v_i, pour\ i = 1, \ldots, n$, on a :

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\cdots+i_n} \times C(x_{1_{i_1}}, \ldots, x_{n_{i_1}}) \leq 0$$

où $x_{1_j} = u_j\ et\ x_{2_j} = v_j, \forall j \in 1, \ldots, n.$

2.2.2 Théorème de Sklar

Théorème 1.21 (Sklar) [5]

Soit F une fonction de répartitions ayant pour marginales $F_1, \ldots, F_n$ alors, il existe une copule C telle que $\forall\ x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, on a :

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \qquad (a)$$

C est unique si $F_1, \ldots, F_n$ sont toutes continues, sinon la copule C est déterminée de manière unique sur $\mathrm{Im}(F_1) \times \cdots \times \mathrm{Im}(F_n)$.

De manière réciproque, si C est une copule et $F_1, \ldots, F_n$ sont des fonctions de répartition univariées alors, F définie au niveau de (a) est la fonction de répartition conjointe de marginale $F_1, \ldots, F_n$.

Corollaire 1.22 Inversion de Sklar [5]

Soient $F, F_1, \ldots, F_n$ et C les fonctions du théorème 1.21 et soient $F_1^{-1}, \ldots, F_n^{-1}$ les inverses généralisés respectifs de $F_1, \ldots, F_n$. Alors,

$$C(u_1, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)) \qquad \forall u \in I$$

2.2.3 Bornes de Fréchet-Hoeffding

Rappelons que les bornes de Fréchet d'une copule multivariée sont données par (voir [1]) :

$$C^-(u_1, \ldots, u_n) = \max\left(\sum_{i=1}^{n} u_n - n + 1, 0\right)$$

et $\qquad C^+(u_1, \ldots, u_n) = \min(u_1, \ldots, u_n).$

On a pour toute copule C

$$W = C^-(u) \leq C(u) \leq C^+(u) = M \qquad \forall u \in I^n.$$

La copule produit associée à $u = (u_1, \ldots, u_n)$ est exprimée comme suit :

$$C^{\perp}(u_1, \ldots, u_n) = \prod_{i=1}^{n} u_i.$$

Il apparaît clairement que des variables aléatoires ayant cette copule sont indépendantes.

2.2.4 Densité d'une copule multivariée

Supposons que la copule C et les distributions marginales $F_1, ..., F_n$ sont différentiables alors, la densité jointe de la variable aléatoire $X = (X_1, ..., X_n)$ prend la forme suivante (voir [1]) :

$$f(x_1, \ldots, x_n) = f_1(x_1) \times \cdots \times f_n(x_n) c(F_1(x_1), \ldots, F_n(x_n)),$$

où pour $f_k, 1 \leq k \leq n$ est la densité de probabilité dérivée de F_k, f est la densité jointe issue de F et c est la densité de la copule C définie par :

$$c(u_1, \ldots, u_n) = \frac{\partial^n}{\partial u_1 \ldots \partial u_n} C(u_1, \ldots, u_n).$$

Nous remarquons qu'on peut séparer la densité jointe en deux blocs.

1. La densité $c(f_1(x_1), \ldots, f_n(x_n))$ contient toute l'information portant sur la structure de dépendance des variables aléatoires $X_1, \ldots, X_n$.

2. Le produit des densités marginales contient l'information sur les marges. Cela prouve que les copules représentent une manière d'extraire la structure de dépendance de la distribution jointe et de la séparer des comportements marginaux.

Grâce à la densité jointe et aux densités marginales, il est alors possible de déterminer la densité d'une copule.

En effet, supposons que la fonction de répartition F admettant la densité continue f alors, les lois marginales $F_1, \ldots, F_n$ sont continues et ont des densités respectives $f_1, \ldots, f_n$; la copule de F admet la densité c définie par :

$$(u_1, \ldots, u_n) = \frac{f(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n))}{f_1(F_1^{-1}(u_1)), \ldots, f_n(F_n^{-1}(u_n))}.$$

Théorème 1.23 [3]

Soient des variables aléatoires continues $X_1, \ldots, X_n$ de copule C.

Si $\alpha_1, \ldots, \alpha_n$ sont des fonctions strictement croissantes sur $\mathrm{Im}(X_1), \ldots, \mathrm{Im}(X_n)$ respectivement, alors $\alpha_1(X_1), \ldots, \alpha_n(X_n)$ ont la même copule C. Ainsi C est invariante par toute transformation strictement croissante des variables aléatoires $X_1, \ldots, X_n$.

Preuve [1]

On note $G_1, \ldots, G_n$ les distributions de $\alpha_1(X_1), \ldots, \alpha_n(X_n)$ respectivement et C_α la copule associée. Comme pour tout k, α_k est strictement croissante, on a :

$$G_k(x) = P[\alpha_k(X_k) \leq x] = P[X_k \leq \alpha_k^{-1}(x)] = F_k(\alpha_k^{-1}(x)).$$

$\forall\, x \in \mathbb{R}$, d'où

$$\begin{aligned} C_\alpha(G_1(x_1),\ldots,G_n(x_n)) &= P[\alpha_1(X_1)\leq x_1,\ldots,\alpha_n(X_n)\leq x_n] \\ &= P[X_1\leq \alpha_1^{-1}(x_1),\ldots,X_n\leq \alpha_n^{-1}(x_n)] \\ &= C(F_1(\alpha_1^{-1}(x_1)),\ldots,F_n(\alpha_n^{-1}(x_n))) \\ &= C(G_1(x_1),\ldots,G_n(x_n)). \end{aligned}$$

Ainsi, on conclut que $C_\alpha = C$.

Théorème 1.24 [3]

Soient des variables aléatoires continues $X_1,\ldots,X_n$ de copule $C_{X_1,\ldots,X_n}$.

Soient $\alpha_1,\ldots,\alpha_n$ des fonctions strictement monotones sur $\mathrm{Im}(X_1),\ldots,\mathrm{Im}(X_n)$ respectivement. Notons $C_{\alpha_1(X_1),\ldots,\alpha_n(X_n)}$ la copule de $\alpha_1(X_1),\ldots,\alpha_n(X_n)$.

Supposons que la fonction α_k est strictement décroissante pour un certain k, par exemple $k=1$, alors, on a :

$$\begin{gathered} C_{\alpha_1(X_1),\ldots,\alpha_n(X_n)}(u_1,\ldots,u_n) = \\ C_{\alpha_2(X_2),\ldots,\alpha_n(X_n)}(u_2,\ldots,u_n) - C_{X_1,\alpha_2(X_2),\ldots,\alpha_n(X_n)}(1-u_1,\ldots,u_n). \end{gathered}$$

2.3 Les copules paramétriques

Dans cette partie, nous allons présenter quelques copules paramétriques les plus utilisées. Nous exposerons en premier lieu les copules Archimédiennes, ensuite nous aborderons la famille des copules Elliptiques. Nous avons considéré le cas bivarié.

2.3.1 Copule Archimédienne

Définition 1.25 [8]

Soit φ une fonction continue, strictement décroissante de I dans $[0,\infty]$ telle que $\varphi(1)=0$ et $\varphi^{[-1]}$ la fonction inverse généralisée de φ. Soit C une fonction définie de I^2 dans I telle que :

$$C(u,v)=\varphi^{[-1]}(\varphi(u)+\varphi(v)).$$

Cette fonction est une Copule si et seulement si φ est convexe. Ce type de copule est appelé Copule Archimedienne, et la fonction φ est appelée générateur de la copule.

Ci-dessous, nous allons donner les propriétés de symétrie et d'associativité des copules Archimediennes.

Théorème 1.26 [4]

Soit une copule Archimédienne C avec un générateur φ alors,

— C est symétrique, ce qui signifie que $C(u,v)=C(v,u)$.

— C est associative, c'est-à-dire, $C(u,C(v,w))=C(C(u,v),w)$, $\forall\ u,v,w\in I$.

Théorème 1.27 [9]

Soient X_1 et X_2 des variables aléatoires dont la copule C est Archimédienne de générateur φ. Alors, le tau de Kendall des variables aléatoires X_1 et X_2 est donné par :

$$\tau=1+4\int_0^1\frac{\varphi(t)}{\varphi^{'}(t)}dt.$$

Notons que le tau de Kendall est une mesure de concordance très utilisée en pratique.

Théorème 1.28 [4]

Soit C une copule Archimédienne alors nous avons :

— Si $(\varphi^{-1})^{'}(0)$ est finie alors,

$$C(u,v)=\varphi^{-1}(\varphi(u)+\varphi(v))$$

n'a pas de dépendance supérieure au niveau de la queue de distribution.

— Si C présente une dépendance supérieure au niveau de la queue de distribution alors, $(\varphi^{-1})^{'}(0)=-\infty$ et de plus on a la relation :

$$\lambda_U=2-2\lim_{s\to 0}\frac{(\varphi^{-1})^{'}(2s)}{(\varphi^{-1})^{'}(s)}.$$

On peut aussi citer un résultat similaire pour les indices de queue inférieurs.

Théorème 1.29 [1]

Soit C une copule Archimédienne. Le coefficient de dépendance de queue inférieure de la copule

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)),$$

est donné par :

$$\lambda_L = 2 \lim_{s \to 0} \frac{(\varphi^{-1})'(2s)}{(\varphi^{-1})'(s)}.$$

La densité d'une copule Archimédienne

La densité d'une copule Archimédienne, de générateur φ deux fois différentiable, est donnée par (voir[4]) :

$$c_\alpha(u,v) = -\frac{\varphi''_X(C_\alpha(u,v))\varphi'_X(u)\varphi'_X(v)}{(\varphi'_X C_\alpha(u,v))^3}.$$

Maintenant, nous allons présenter les familles de copule Archimédienne les plus célèbres.

A) La famille de Frank [2]

L'expression du générateur de la copule de Frank est la suivante :

$$\varphi(t) = -\ln\left(\frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}\right) \quad \text{avec } \alpha \neq 0 \text{ et } t \in]0,1].$$

Nous obtenons ainsi la copule de Frank qui est définie de la manière suivante :

$$C_\alpha(u,v) = -\frac{1}{\alpha}\ln\left(1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1}\right).$$

Le tau de Kendall

Le tau de Kendall correspondant à cette famille est donné par la formule suivante :

$$\forall\ k \geq 0, \qquad \tau_\alpha = 1 - \frac{4}{\alpha}(1 - D_k(x)) \qquad \text{avec} \quad D_k(x) = \tfrac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt.$$

La famille de Frank ne possède pas de dépendance de queue supérieure et inférieure.

B) La famille de Clayton [2]

Concernant la copule de Clayton, son générateur s'écrit de la manière suivante :

$$\text{Pour } \alpha > 0 \text{ et } v \in]0, 1], \qquad \varphi(v) = \alpha^{-1}(v^{-\alpha} - 1).$$

On obtient alors la copule de Clayton (en dimension deux) :

$$C_\alpha(v_1, v_2) = (v_1^{-\alpha} + v_2^{-\alpha} - 1)^{-1/\alpha}.$$

Le tau de Kendall

Le tau de Kendall pour la copule de Clayton est :

$$\tau(\alpha) = \frac{\alpha}{\alpha + 2}.$$

C) La famille de Gumbel [8]

Le générateur de la copule de Gunbel est donné par :

$$\varphi(t) = (-\ln(t))^\alpha \qquad \alpha \geq 1 \text{ et } t \in [0, 1].$$

La fonction copule associée est donnée par l'expression ci-après :

$$C_\alpha(u, v) = \exp(-[(-\ln(u))^\alpha + (\ln(v))^\alpha]^{1/\alpha}) \qquad \forall\, \alpha \geq 1,\ \forall (u, v) \in I^2.$$

Le tau de Kendall

Le tau de Kendall pour la famille de Gumbel est défini par :

$$\tau(\alpha) = 1 - \tfrac{1}{\alpha} \ \text{ avec } \alpha \geq 1.$$

Il est à noter que la copule de Gumbel s'adapte mieux aux données qui ont une forte dépendance à droite et une faible dépendance à gauche.

2.3.2 Copule Elliptique

Les copules elliptiques sont définies à partir des lois de distribution elliptique. A ce niveau, nous donnons quelques définitions de la distribution elliptique.

Définition 1.30 [6]

Un vecteur aléatoire $X = (X_1, \ldots, X_n)$ suit une distribution elliptique s'il a une représentation décrite par la relation suivante :

$$X = \mu + RAU,$$

où,

— $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$,
— U est un vecteur aléatoire uniforme sur la sphère unité de $\mathbb{R}^n$
— R est un vecteur aléatoire indépendant de U,
— A est une matrice de dimension $n \times n$ telle que $\Sigma = AA^T$ est non singulière.

Définition 1.31 [8]

La fonction de densité d'une distribution elliptique si elle existe est donnée par :

$$f(x) = |\Sigma|^{-1/2} g((x-\mu)^\top \sum^{-1}(x-\mu)), x \in \mathbb{R}^n,$$

où g est une fonction définie de $\mathbb{R}^+$ dans $\mathbb{R}$ et $|\Sigma|$ est le déterminant de Σ.

Définition 1.32 [10] [8]

Soit un vecteur aléatoire de distribution elliptique $X = (X_1, \ldots, X_n)$.

La fonction caractéristique $\phi_X(t), t \in \mathbb{R}^n$ est donnée par :

$$\begin{aligned} \phi_X(t) &= E(\exp\ it^\top X) \\ &= E(\exp(it^T\mu + RAU)) \\ &= \exp(it^\top\mu) g(t^\top \Sigma t). \end{aligned}$$

On note par $X \sim E_n(\mu, \Sigma, g)$ la classe de la distribution elliptique de vecteur moyenne μ, de matrice de covariance $\Sigma = (\delta_{ij}), 1 \leq i, j \leq n$ et de générateur caractéristique g.

La copule gaussienne fait partie des exemples de copule elliptique.

A) La copule Gaussienne bivariée

Définition 1.33 [2]

La copule Gaussienne ou copule normale bivariée est définie comme suit :

$$C(u_1, u_2; \rho) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)).$$

— ρ est le coefficient de corrélation,

— Φ_p la distribution normale bivariée standard dont la matrice de corrélation est :

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

On déduit alors la densité de la copule Gaussienne bidimensionnelle en dérivant l'expression qui définit la copule Gaussienne :

$$C(u_1, u_2) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(- \frac{\beta_1^2 + \beta_2^2 - 2\rho\beta_1\beta_2}{2(1-\rho^2)} + \frac{\beta_1^2 + \beta_2^2}{2} \right)$$

$\rho \in (-1, 1)$

Avec $\beta_1 = \Phi^{-1}(u_1)$ et $\beta_2 = \Phi^{-1}(u_2)$.

B. Les propriétés de la copule Gaussienne

Tau de Kendall et Rho de Spearman[8]

Le tau de Kendall d'une copule Gaussienne est donné par :

$$\tau = \frac{2}{\pi} \arcsin(\rho^{'}),$$

et le rho de Spearman est donné par :

$$\rho = \frac{6}{\pi} \arcsin\left(\frac{\rho^{'}}{2}\right).$$

$\rho^{'}$ représente le paramètre de la copule.

Un autre exemple de copules elliptiques est la copule de Student.

C. La copule de Student bivariée

Définition 1.34 [11]

La copule de Student bivariée est définie comme suit :

$$C(u, v; \rho, \nu) = t_{\rho,\nu}(t_{\nu}^{-1}(u), t_{\nu}^{-1}(v)),$$

avec ρ désigne le coefficient de corrélation de Pearson et t_ν représente la fonction de répartition de la loi de Student avec ν degré de liberté.

La copule de Student bivariée est alors donnée par :

$$C_{\nu,\rho}(u, v) = \int_{-\infty}^{t_{\nu}^{-1}(u)} \int_{-\infty}^{t_{\nu}^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{\beta_1^2 + \beta_2^2 - 2\rho\beta_1\beta_2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} d\beta_1 d\beta_2$$

avec $0 < \rho < 1$.

La densité de la copule de Student s'écrit par :

$$c_{\nu,\rho}(u) = \rho^{-1/2} \frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})} \left(1 + \frac{\beta_1^2 + \beta_2^2 - 2\rho\beta_1\beta_2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} \times \left(1 + \frac{\beta_1}{\nu}\right)^{-\frac{\nu+2}{2}} \left(1 + \frac{\beta_2}{\nu}\right)^{-\frac{\nu+2}{2}}.$$

D. Les propriétés de la copule Student

Tau de Kendall et Rho de Spearman

Le tau de Kendall de la copule Student est identique au tau de Kendall de la copule Gaussienne.

2.4 Estimation des copules

En statistique, l'estimation est un processus qui a pour rôle de déterminer un paramètre α inconnu du modèle statistique à partir d'un échantillon observé $(X_1, \ldots, X_n)$.

Il existe des méthodes d'estimation paramétriques telle que la méthode des moments, la méthode du maximum de vraisemblance. Il existe aussi des méthodes d'estimation non-paramétrique, comme les méthodes basées sur la copule emipirique, par exemple.

2.4.1 Estimation paramétrique

Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire avec des fonctions de répartition marginale univariée $F_j(x_j, \delta_j),\ \ j = 1, \ldots, n$.

Soit C une copule appartenant a une famille paramétrique $C = \{C_\theta, \theta \in \Theta\}$. La fonction de répartition de X est donnée par (en considérant le théorème de Sklar) :

$$F(x_1, \ldots, x_n) = C\{F_1(x_1; \delta_1), \ldots, F_n(x_n; \delta_n); \theta\},$$

et sa densité par :

$$f(x_1, \ldots, x_n; \delta_1, \ldots, \delta_n, \theta) = c\{F_1(x_1; \delta_1), \ldots, F_n(x_n; \delta_n); \theta\} \prod_{j=1}^{n} f_j(x_j; \delta_j),$$

avec

$$c(u_1, \ldots, u_n) = \frac{\partial^n C(u_1, \ldots, u_n)}{\partial u_1 \cdots \partial u_n}.$$

Pour un échantillon d'observation $\{x_k\}_{k=1}^{K},\ x_k = (x_1^k, \ldots, x_n^k)^T$ et le vecteur des paramètres $\alpha = (\delta_1, \ldots, \delta_n, \theta) \in \mathbb{R}^{n+1}$, la fonction de vaisemblance est donnée par :

$$L(\alpha; x_1, \ldots, x_k) = \prod_{k=1}^{K} f(x_1^k, \ldots, x_n^k; \delta_1, \ldots, \delta_n, \theta).$$

La fonction de log-vraisemblance est donnée par :

$$l(\alpha; x_1, \ldots, x_k) = \sum_{k=1}^{K} \log c\{F_1(x_1^k; \delta_1), \ldots, F_n(x_1^k; \delta_n); \theta)\} + \sum_{k=1}^{K} \sum_{j=1}^{n} \log f_j(x_j^k; \delta_j).$$

A. Maximum de vraisemblance (MLE)

Définition 1.35 [12]

Pour l'estimation en basant sur la méthode de maximum de vraisemblance, le vecteur des paramètres α est estimé par une seule étape :

$$\widehat{\alpha} = \arg\max\, l(\alpha).$$

B. Inférence sur les marginales (IMF)[8]

La méthode de l'inférence repose sur le fait qu'il faut estimer séparément les paramètres de la copule et les marges.

Ainsi, la log-vraisemblance peut s'écrire comme suit :

$$l(\delta) = \sum_{k=1}^{K} \log c\{F_1(x_1^k;\delta_1), ..., F_n(x_n^k;\delta_n);\theta)\} + \sum_{k=1}^{K}\sum_{j=1}^{n} \log f_j(x_j^k;\delta),$$

où $\delta = (\delta_1, ..., \delta_n, \theta)$.

La méthode de l'inférence sur les marginales s'effectue en deux étapes :

1. On estime les paramètres des lois marginales

$$\begin{aligned}\widehat{\delta}_j &= \arg\max\ l_j(\delta_j) \\ &= \arg\max\ \sum_{k=1}^{K} \log f_j(x_j^k;\delta_k).\end{aligned}$$

2. Ensuite on estime θ en effectuant la maximisation de la fonction suivante :

$$l(\theta, \widehat{\delta}_1, \ldots, \widehat{\delta}_n) = \sum_{k=1}^{K} \log c\{F_1(x_1^k;\widehat{\delta}_1), \ldots, F_n(x_n^k;\widehat{\delta}_n);\theta)\}.$$

Le paramètre de dépendance estimer $\widehat{\theta}$ est donné par :

$$\widehat{\alpha}_{IFM} = (\widehat{\delta}_1, \ldots, \widehat{\delta}_n, \widehat{\theta})^\top.$$

C. La méthode des moments [5]

En utilisant, la méthode des moments, les paramètres θ_i, $i = 1, \ldots, d$ des lois marginales et le paramètre α de la copule sont estimé. La démarche est la suivante :

1. Résoudre le système :

$$\begin{cases} \overline{X_n} = f(\theta_1, \ldots, \theta_d) \\ S_n^2 \ = g(\theta_1, \ldots, \theta_d) \\ \mu_{3,n} = h(\theta_1, \ldots, \theta_d) \\ \quad \vdots \end{cases} \tag{2.1}$$

où d représente la dimension de θ, f, g et h sont les expressions des moments d'ordre 1, 2 et 3 en fonction du paramètre θ_i. Répéter cette étape pour toutes les marginales.

2. Pour obtenir le paramètre α de la copule, il suffit d'inverser le tau de Kendall ou le rho de Spearman.

Exemple :

Pour la copule de Gumbel de paramètre θ, on a :

$$\tau = 1 - \frac{1}{\theta}.$$

Nous concluons que :

$$\theta = \frac{1}{1-\tau}.$$

Une estimation du paramètre de la copule obtenue en posant $\widehat{\theta} = \dfrac{1}{1-\widehat{\tau}}$ si nous possèdons une estimation $\widehat{\tau}$ du tau de Kendall :

En général, l'estimateur non paramétrique du tau de Kendall est donné par l'expression :

$$\widehat{\tau} = \frac{c-d}{c+d},$$

où c et d représentent respectivement le nombre de paires disjointes concordantes et disconcordantes.

2.4.2 Estimation semi paramétrique

Maximum de vraisemblance canonique [8]

Au niveau de la méthode de maximum de vraisemblance canonique, la distribution marginale univariée est estimée en partant de la fonction empirique $\widehat{F}$ pour $j = 1, \ldots, n$, (voir[12])

$$\widehat{F}_j(x) = \frac{1}{K+1} \sum_{k=1}^{K} \mathbb{I}(x_{j,k} \leq x),$$

où $j \in 1, ..., d$ et $\mathbb{I}$ est la fonction indicatrice.

La fonction de log-vraisemblance est :

$$l(\theta) = \sum_{k=1}^{K} \log c(\{\widehat{F}_1(x_{1,k}), \ldots, \widehat{F}_n(x_n^k); \theta\}),$$

et le paramètre de la copule estimer $\widehat{\theta}_{CML}$ est donnée par :

$$\widehat{\theta}_{CML} = \arg\max\, l(\theta).$$

2.4.3 Estimation non paramétrique

Définition 1.36 [5]

Soit un échantillon $(X_1, \ldots, X_n)$ de taille n de loi F. Si nous avons $(x_1^k, \ldots, x_n^k), 1 \leq k \leq K$ un K-échantillon du vecteur X (de dimension n), on peut généraliser l'expression de la fonction de répartition empirique :

$$F_n(x_1, \ldots, x_n) = \frac{1}{K} \sum_{i=1}^{K} \mathbb{I}(X_1^k \leq x_1, \ldots, X_n^k \leq x_n).$$

Définition 1.37 [5]

Soit un échantillon $\{(x_k, y_k)\}_{k=1}^n$ de taille n d'un copule de variables aléatoires.

La copule empirique de la fonction de répartition est donnée par :

$$\widehat{C}\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\text{Nombre des paires } (x, y) \text{ dans l'échantillon tels que } x \leq x_{(i)} \text{ et } y \leq y_{(j)}}{n}$$

où $x_{(i)}$ et $y_{(j)}$ sont les statistiques d'ordre associées à l'échantillon.

La fonction densité empirique notée $\widehat{c}$ de la copule C est donnée par :

$$\widehat{c}\left(\frac{i}{n},\frac{j}{n}\right)=\begin{cases}\dfrac{1}{n}\text{ si }(x_{(i)},y_{(j)})\text{ est un élément de l'échantillon}\\ 0\text{ si non}\end{cases}$$

Elle est parfois appelée **fréquence empirique de la copule**. La relation entre $\widehat{C}$ et $\widehat{c}$ est la suivante :

$$\widehat{C}\left(\frac{i}{n},\frac{j}{n}\right)=\sum_{p=1}^{i}\sum_{q=1}^{j}\widehat{c}\left(\frac{p}{n},\frac{q}{n}\right),$$

et

$$\widehat{c}\left(\frac{i}{n},\frac{j}{n}\right)=\widehat{C}\left(\frac{i}{n},\frac{j}{n}\right)-\widehat{C}\left(\frac{i-1}{n},\frac{j}{n}\right)-\widehat{C}\left(\frac{i}{n},\frac{j-1}{n}\right)+\widehat{C}\left(\frac{i-1}{n},\frac{j-1}{n}\right).$$

Définition 1.38 [5]

Soit un échantillon $(x_k^1,\ldots,x_k^n)_{k=1}^K$, de taille n et de dimension K.

Dans le cas multivarié, la copule empirique est donnée par :

$$\widehat{C}\left(\frac{k_1}{K},\ldots,\frac{k_n}{K}\right)=\begin{cases}\dfrac{1}{K}\ si\ (x_k^1\leq x_{(k_1)}^1,\ldots,x_{(k)}^n\leq x_{(k_n)}^n),\\ 0\text{ si non.}\end{cases}$$

Les copules empiriques peuvent être utilisées pour estimer les mesures de dépendance. On peut ainsi proposer l'estimateur du rho de Spearman et du tau de Kendall dans le cas bivarié :

$$\widehat{\rho}=\frac{12}{n^2-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\widehat{c}\left(\frac{i}{n},\frac{j}{n}\right)-\frac{ij}{n^2}\right),$$

$$\widehat{\tau}=\frac{2n}{n-1}\sum_{i=2}^{n}\sum_{j=2}^{n}\sum_{p=1}^{i-1}\sum_{q=1}^{j-1}\left(\widehat{c}\left(\frac{i}{n},\frac{j}{n}\right)\widehat{c}\left(\frac{p}{n},\frac{q}{n}\right)-\widehat{c}\left(\frac{i}{n},\frac{q}{n}\right)\widehat{c}\left(\frac{p}{n},\frac{j}{n}\right)\right).$$

Voir [4], pour la démonstration.

2.5 Conclusion

Dans ce chapitre, plusieurs définitions, théorèmes et démonstrations sont proposés sur la notion relative aux copules. Dans le chapitre suivant, nous rappellerons quelques généralités sur les réseaux V2G et les systèmes de détection d'intrusions (SDI).

Chapitre 3

GÉNÉRALITÉS SUR LE V2G

Le V2G (Vehicle-to-grid) est une technologie qui permet aux véhicules électriques (les véhicules électriques à batterie, les véhicules hybrides rechargeables) de se recharger ou de distribuer une partie de leur énergie électrique au réseau électrique. L'énergie électrique qui transite à travers les réseaux V2G est issue de sources d'énergie renouvelables telles que le solaire et l'éolienne. Grâce à cette technologie, on assiste à la naissance d'un autre moyen de se faire de l'argent et les voitures deviennent moins polluantes (moins émettrices de CO_2).

D'après la norme ISO 15118-1, les entités du réseau véhiculaire V2G sont classifiées en deux grandes catégories. La première catégorie prend en compte les acteurs primaires et l'autre catégorie les acteurs secondaires.[13]

Les acteurs primaires sont directement impliqués dans tous les services rendus par le réseau sus-cité et sont énumérés comme suit :

— le véhicule électrique qui est composé d'une batterie, d'un contrôleur de communication, d'une interface homme-machine, et d'une unité de contrôle électronique ;

— la borne de recharge qui est constituée d'un contrôleur de communication, d'une unité de paiement, d'une interface homme-machine, et d'un compteur électrique.

Il est à noter que les acteurs cités ci-dessus possèdent chacun un contrôleur de communication. Le contrôleur de communication est essentiel pour établir, maintenir et rompre une communication entre ces derniers.

Concernant les acteurs secondaires, ils interviennent indirectement dans le processus d'échanges et au nombre de ceux-ci on trouve : l'opérateur de mobilité, le fournisseur électrique et le constructeur automobile.

Comme tout système informatique, le réseau V2G est confronté à plusieurs types d'attaques. Au nombre de ces attaques, on peut citer :

— Le déni de service (consiste à rendre indisponible une ressource matérielle ou logicielle aux clients par la surcharge du réseau en requêtes) ;
— L'usurpation d'identité (consiste à falsifier la communication entre un véhicule électrique et sa borne de recharge afin de se faire passer pour l'un ou l'autre) ;
— L'attaque par rebond (consiste à attaquer une voiture électrique ou une borne de recharge par l'intermédiaire d'une autre machine afin de masquer ses traces).

Pour protéger le réseau V2G de ces attaques, des SDI et des architectures sécurisées (comme l'Infrastructure à Clés Publiques) ont été proposés par différents chercheurs.

Les systèmes de détection d'intrusions ont pour rôle de détecter des activités suspectes ou anormales sur la base des données collectées dans un réseau(il s'agit ici du réseau V2G). Ces systèmes sont classifiés généralement en deux (02) catégories à savoir :

— Les SDI basés sur la signature ;

  Avec ce type de SDI, il est important d'avoir une base de données des signatures d'attaques à jour afin de ne pas exposer son réseau à des attaques.

  Les SDI basés sur la signature ne protègent pas des attaques non connues ou non répertoriées.
— Les SDI basés sur les anormalies ;

  Il s'agit de la capacité du système de détection d'intrusions à différencier un comportement anormal d'un comportement normal. Afin d'avoir cette capacité, il faudrait que le SDI soit préalablement entrainé. Pour effectuer l'entrainement d'un SDI, plusieurs techniques sont mises en œuvre comme l'apprentissage artificiel, la technologie blockchain.

Dans le présent travail, notre modèle de prédiction d'attaques est relatif à un SDI basé sur les anormalies.

Toutefois, il est possible de rencontrer des systèmes de détection d'intrusions hybrides qui représentent une combinaison des deux (02) types.

3.1 Conclusion

De ce chapitre, on retient que les réseaux V2G aussi ne sont pas à l'abri des attaques informatiques. Il est important de les protéger en proposant des architectures V2G sécurisées avec des SDI performants. Dans le chapitre suivant, nous présentons un état de l'art sur les objectifs de sécurité des réseaux V2G et quelques solutions pour détecter les attaques dans plusieurs réseaux (V2G, VANET,...).

Chapitre 4

REVUE DE LA LITTÉRATURE

D'après l'United States Environmental Protection Agency (EPA), les principales sources d'émission de gaz à effet de serre aux Etats-Unis sont : le transport (29 %), la production d'électricité (25 %), l'industrie (23 %), le commercial et le résidentiel (13 %), l'agriculture (10 %), l'utilisation des terres et foresterie (12 %)[14]. Suite à ce constat, les dirigeants américains ont prévu un déploiement massif d'infrastructure de recharge afin de réduire la quantité de CO_2 émise par les véhicules. Leur objectif est d'atteindre 2,4 millions de bornes recharges d'ici 2030[15].

Dans le cadre du déploiement des bornes de recharge, il est important de connaître les exigences de sécurité du réseau V2G et de mettre en place les contrôles de sécurité nécessaires pour minimiser le risque de compromission des entités du réseau par les cybercriminels.

Dans ce chapitre, nous présentons d'une part les objectifs en matière de sécurité du réseau V2G puis d'autre part les solutions proposées par les chercheurs pour détecter et déjouer les attaques dans plusieurs réseaux.

4.1 Défis de sécurité dans les réseaux V2G

La question de la sécurité constitue une problématique majeure dans les réseaux véhiculaires et plus précisément dans le réseau V2G. L'Union Internationale des Télécommunications (UIT) conformément à ses prérogatives a pris une recommandation au

sujet des lignes directrices relatives à la sécurité des communications de véhicule à tout autre élément (V2X). Il s'agit de la recommandation UIT-T X.1372 [16]. Dans cette recommandation, il est question des menaces observées dans le cadre du déploiement des systèmes de communications V2X et des exigences en matière de sécurité pour ce type de communications. Aussi, elle donne des directives concernant la mise en œuvre de communications V2X sécurisées afin de répondre aux exigences de sécurité.

Les attaques informatiques sont perçues comme des failles d'un réseau exploitées par les hackers dans le but d'effectuer des activités malveillantes comme voler des informations à caractère personnel ou confidentiel, troubler le bon fonctionnement d'un service et bien d'autres. D'après la recommandation [16], un réseau V2X (qui prend en compte le réseau V2G) peut être confronté aux attaques suivantes : le déni de service, l'usurpation d'identité, l'attaque temporelle, l'attaque de l'homme du milieu. De cette même recommandation, nous pouvons déduire que les exigences relatives à la sécurité du V2G sont : la confidentialité, l'intégrité, la disponibilité, la non-répudiation, l'authenticité et l'imputabilité.

Dans [17], les notions relatives à la technologie V2G ont été abordées. Il s'agit notamment des concepts de base, des avantages, de l'architecture et de la cybersécurité dans le V2G. Les objectifs de sécurité dans les réseaux V2G sont : la confidentialité, l'intégrité, la disponibilité et l'authenticité. Face au risque de piratage, un modèle basé sur la cyber-assurance est proposé. L'objectif de ce modèle est de transférer le risque de compromission d'un véhicule électrique à un assureur.

Dans [18], les exigences en matière de sécurité du réseau V2G sont : la confidentialité, l'authenticité et l'intégrité. Pour l'atteinte de ces objectifs, les auteurs ont proposé une architecture sécurisée de charge/décharge des véhicules électriques avec des mesures de sécurité comme l'authentification anonyme, le contrôle d'accès, la signature anonyme et l'attestation à distance.

Les questions relatives aux échanges bidirectionnels sécurisés entre un véhicule électrique (EV) et une borne de recharge (EVSE) sont traitées par la norme ISO 15118. Cette norme se présente sous plusieurs versions parmi lesquelles on trouve :

— ISO 15118-1 : aborde les généralités et les cas d'utilisation du standard ISO

15118

— ISO 15118-2 : traite des exigences liées aux protocoles des couches 3 à 7 du modèle OSI

— ISO 15118-3 : traite des exigences liées aux protocoles des couches 1 et 2 du modèle OSI.

Toutefois, il est important de relever que la norme ISO 15118 bien qu'étant une référence dans le domaine de la sécurité du V2G, celle-ci ne prend pas en compte tous les paramètres liés à la sécurité. Ses insuffisances sont corrigées par les travaux des chercheurs. Plusieurs solutions innovantes sont proposées dans la littérature. Elles permettent de renforcer la sécurité des échanges de données entre un véhicule électrique et une borne de recharge. Parmi les solutions proposées, les SDI occupent une place de choix.

4.2 Détection d'attaques sur les réseaux V2G

Dans [19], les auteurs ont effectué une enquête sur les techniques de détection d'intrusions dans un réseau. D'après cet article, l'apprentissage automatique, l'apprentissage en profondeur et la technologie du blockchain sont largement utilisés en cybersécurité pour détecter les anomalies dans les réseaux. L'apprentissage automatique et l'apprentissage en profondeur s'appuient sur des algorithmes spécifiques afin de mettre en lumière toute intrusion. Par exemple, l'apprentissage automatique se base sur les algorithmes tels que le Support Vector Machine (SVM), le classificateur Bayes, la table de décision, l'arbre de décision, le clustering, le k-means tandis que l'apprentissage en profondeur s'appuie sur les algorithmes comme l'Auto Encoder, le Convolutional Neural Network (CNN), le Deep Neural Network (DNN), le Recurrent Neural Network (RNN), le Boltzmann Machine (BM).

Dans [20], les auteurs ont proposé un modèle de détection d'intrusions basé sur un réseau de neurone profond. Ce modèle est destiné au réseau Internet. La méthodologie utilisée pour la réalisation du modèle se résume à quatre étapes. La première étape consiste à télécharger d'une base de données nommée CICIDS2017 qui contient les

attaques informatiques les plus récentes durant la période de recherche. Ensuite cette base de données a subit des opérations de nettoyage dans le but de réduire le nombre de variables ou fonctionnalités (passage de 83 variables à 44 variables). Suite à cette étape, la base de données nettoyée a été standardisée à cause de la présence de données asymétriques. La dernière étape consiste à fournir les 44 variables (données nettoyées et standardisées) à un réseau de neurones profonds. Le réseau de neurones profonds utilisé est constitué d'une couche d'entrée suivie de 8 couches composées respectivement de 140, 120, 100, 80, 60, 40, 20 et 120 nœuds et d'une couche finale pour produire les probabilités. Après la réalisation du modèle, sa faculté à prédire une intrusion dans un réseau a été évaluée en se basant sur les métriques ci-après : precision, recall et F-measure. Il ressort de cette évaluation que le modele proposé a une precision de 94,31%, un recall de 95,62 % et une F-measure de 94,1 %.

Dans [21], il est proposé un système d'inférence neuro-floue adaptatif (ANFIS en anglais) dont le but est de prédire un indicateur de sécurité dans les réseaux VANET. Le modèle ANFIS proposé est un système qui utilise un réseau de neurones s'appuyant sur un système d'inférence floue de type Takagi Sugeno. La méthodologie utilisée pour la réalisation du modèle se résume en cinq étapes. La première étape consiste à générer une base de données suite à différentes simulations effectuées à l'aide du simulateur VEINS (Vehicles in Network Simulation). Les simulations découlent de deux scénarios : un scénario sans attaque et l'autre avec des attaques de type DoS. La deuxième étape est la phase de sélection des variables qui est suivi de l'étape étude de corrélation. Le but de cette troisième étape est d'analyser les relations entre les variables sélectionnées. L'étape suivante est l'utilisation de l'Analyse en Composantes Principales (ACP) afin de détecter les variables les plus significatives pour éviter la redondance des données. Enfin, le système ANFIS a été défini et construit à l'aide de MATLAB Toolbox. Le modèle proposé a été évalué sur sa capacité à prédire une attaque. De cette évaluation, on retient que le modèle arrive à prédire les véhicules non attaqués mais la prédiction est moins bonne pour les véhicules attaqués. Cependant la marge d'erreur est acceptable pour garantir l'efficacité du modèle.

Dans le but de déjouer les attaques dans les réseaux VANET, il est proposé dans [22] un modèle basé sur la notion de Watchdog et la théorie des réseaux bayésiens.

La méthode Watchdog consiste à repérer les nœuds qui effectuent des activités malveillantes. En effet, les nœuds avec un Watchdog implémenté ont pour rôle d'écouter en continu les nœuds voisins et de surveiller leur comportement. Avec ladite méthode, on se retrouve avec un taux élevé de faux positifs due à la mobilité des nœuds et au bruit. Afin de réduire le taux élevé de faux positifs, un filtre bayésien est utilisé.

Pour protéger les réseaux VANET contre les attaques de type DoS, les auteurs ont étudié plusieurs méthodes [23]. Il s'agit des méthodes comme : Detecting Jamming attacks in Vehicle Ad hoc Network (DJVAN), Request Response Detection Algorithm (PRDA), Attacked Packet Detection Algorithm (APDA), IP-CHOCK, une méthode d'authentification basée sur la signature. L'étude nous permet de retenir que malgré les différentes solutions proposées pour contrecarrer les attaques de type DoS, il n'est actuellement pas possible de faire disparaître les attaques DoS. La seule chose possible est de réduire le niveau de gravité.

Pour lutter contre les attaques floues et DoS, dans [24] il est proposé un système de détection d'intrusions qui s'appuie sur l'apprentissage automatique. Dans le cadre de la réalisation du SDI, les auteurs ont utilisé deux types de données. Il s'agit notamment des « données d'attaque floue » et des « données d'attaque DoS » qui sont fournies par le Hacking and Countermeasure Research Lab (HCRL). Sur ces données, des opérations de nettoyage ont été réalisées et elles ont été suivies d'opérations de normalisation. Ensuite, les données ont été étiquetées afin de différencier les données normales des données injectées. Suite à l'étiquetage des données, les algorithmes SVM et KNN sont construits pour la classification. Dans le cadre de l'évaluation du SDI, une comparaison a eu lieu entre les algorithmes SVM et KNN. On retient de cette comparaison que les deux algorithmes ont donné d'excellents résultats, néanmoins l'algorithme SVM est moins efficace que l'algorithme KNN.

Dans [25], pour protéger les réseaux VANET des attaques de type Sybil, une solution est proposée. L'attaque de Sybil consiste à entourer un véhicule légitime de faux véhicules. La solution proposée pour déjouer cette attaque repose sur la technologie appelée « capteur du système d'aide à la conduite avancé (ADAS en anglais) » qui permet à une voiture de maîtriser son environnement. Le mécanisme de détection proposé

se déroule en trois étapes consécutives. Tout d'abord, nous avons l'étape de validation des messages qui permet de s'assurer que le message reçu provient d'une source valide. Ensuite, il faut vérifier les informations relatives à l'expéditeur et son emplacement et enfin il faut vérifier les objets environnants. De l'analyse des résultats obtenus à partir du simulateur CARLA, il ressort qu'un véhicule est capable de détecter de faux véhicules sur une distance réduite. Sur une longue distance, le nombre de faux positifs est élevé.

En dehors des différentes méthodes sus-citées sur lesquelles reposent les systèmes de détection d'intrusions, on retrouve la méthode mathématique. Les SDI peuvent peuvent s'appuyer sur des modèles mathématiques. C'est à ce titre que dans [26], les auteurs ont proposé un modèle de détection des anomalies fondé sur les copules. Ce modèle est réalisé suivant les étapes ci-après : réalisation de l'arbre d'informations maximales, modélisation des distributions conjointes des paires de variables obtenues par l'arbre d'information maximal et attribution des scores d'anomalie. L'utilisation des données issues des stations de base d'un réseau LTE d'Europe occidentale a permis l'évaluation du modèle proposé. Les résultats qui découlent de cette évaluation sont très satisfaisants. A titre d'exemple, le modèle de détection des anomalies basé sur les copules a donné de meilleurs résultats comparé à d'autres méthodes comme :

— la méthode de détection des anomalies basée sur l'ACP ;
— la méthode de détection des anomalies basée sur la forêt d'isolement ;
— la méthode de détection des anomalies basée sur le facteur aberrant local (local outlier factor en anglais) ;
— la méthode de détection des anomalies basée sur le clustering-LOF ;
— la méthode de détection des anomalies basée sur l'auto-encodeur ;

Dans [27], les auteurs ont proposé différentes méthodes mathématiques afin de contrecarrer les attaques de type DoS dans les réseaux VANET. Il s'agit notamment de "Root Mean Square (RMS)", "Means Absolute Value (MAV) methods", "Mean Squared Error (MSE)" et "Logistic Regression Model". Ils ont également proposé un modèle basé sur un réseau de neurones dédié à la protection des réseaux VANET contre les attaques de type DoS. Les différentes méthodes citées précédemment ont été appliquées à une

base de données résultante de la collecte d'informations relatives aux simulations réalisées à l'aide des outils OMNET++ et SUMO. Avant l'utilisation de la base de données, des opérations de nettoyage ont été effectués. L'analyse des résultats obtenus, montre que la régression logistique et le réseau de neurones sont plus performants que le MSE, RMS, MAV dans le cadre de la prédiction des attaques de type DoS dans les réseaux VANET.

4.3 Conclusion

Comme on a pu le voir dans notre revue de la littérature, plusieurs approches ont été présentées pour réaliser des systèmes de détection d'intrusions (modèles basés sur les réseaux de neurones, l'apprentissage en profondeur ou les mathématiques), certaines donnent de bons résultats (modèles basés sur les réseaux de neurones, l'apprentissage en profondeur ou la régression logistique) et d'autres ont des résultats moins bons (système d'inférence neuro-floue adaptatif et les modèles basés sur les méthodes mathématiques MSE, RMS, MAV). Dans le chapitre suivant, nous présenterons notre modèle de prédiction d'attaques sur les réseaux V2G. Notre modèle sera basé sur la notion de copules.

Chapitre 5

MÉTHODOLOGIE PROPOSÉE

Pour réaliser notre modèle basé sur les copules, une base de données est d'abord générée à partir de trois scénarios (scénario sans attaque, scénario avec attaques de type MitM et scénario avec attaques de type DoS) à l'aide d'outils tels que MiniV2G, Wireshark et CICflowMeter. Ensuite, les variables significatives de la base de données obtenue sont sélectionnées à l'aide de l'ACP. Enfin, l'algorithme de classification basé sur la notion des copules est construit sous le logiciel R.

5.1 Génération de la base de données

Dans le cadre de notre recherche, nous avons généré la base de données d'attaques V2G sous format CSV grâce au simulateur MiniV2G[28]. Elle a été obtenue à partir de trois différents scénarios :

— scénario sans attaque ;

Pour ce scénario cinq entités sont crées : deux véhicules électriques, une borne de recharge, un switch (s1) et un contrôleur(c0). L'architecture de ce scénario se présente comme celle représentée dans la figure 4.1 ci-dessous. Les trois (03) interfaces créées par le switch sont sniffées par l'outil Wireshark lors du processus de charge des véhicules électriques. Afin d'avoir assez de paquets à analyser, le processus de charge a été répété plusieurs fois pour les deux véhicules. Après le sniffing des interfaces du switch, les données capturées sont enregistrées sous format PCAP.

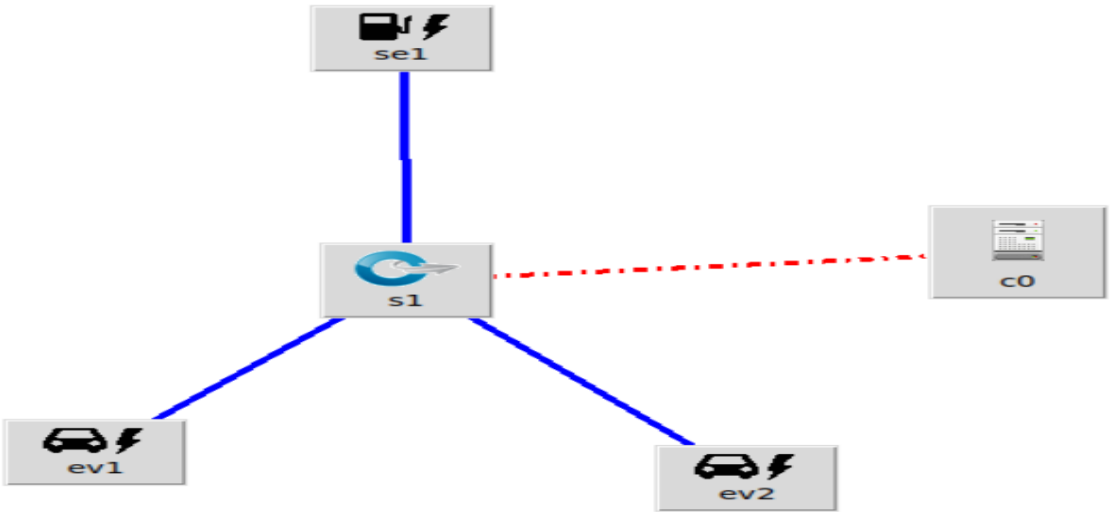

FIGURE 5.1 – Architecture relative à la simulation sans attaque

— scénario avec une attaque de type homme du milieu ;

A ce niveau, quatre entités sont créées : un véhicule électrique, une borne de recharge, un switch (MitM Switch) et une entité MitM. L'entité MitM Switch a pour but d'intercepter le flux de données quittant le EVCC et de le rediriger vers l'entité MitM. Ainsi, le noeud MitM est en mesure de faire toute sorte de modification avant l'envoi du flux de données vers le SECC. La figure 4.2 ci-dessous décrit clairement l'architecture du présent scénario. Lors du processus de charge du véhicule électrique, il y a eu un sniffing des trois interfaces du MitM Switch grâce à l'outil Wireshark. Suite à la capture, les données capturées sont enregistrées sous format PCAP.

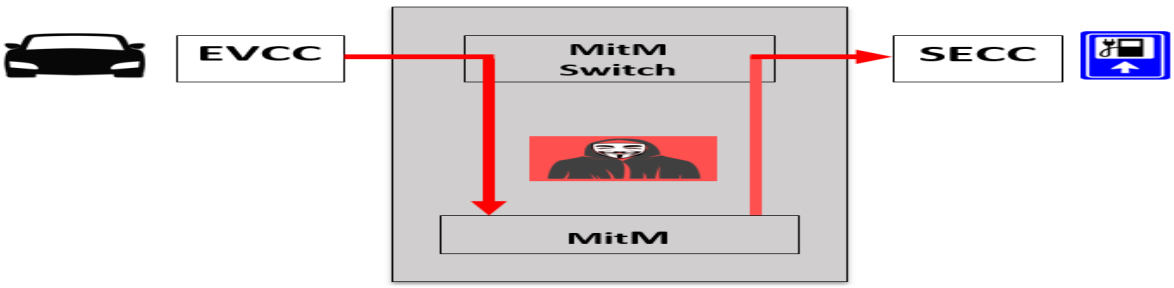

FIGURE 5.2 – Architecture relative à la simulation de l'attaque de type MitM

— scénario avec une attaque de type DoS ;

Dans ce scénario, nous avons aussi quatre entités qui sont créées comme dans

le scénaro précédent : un véhicule électrique, une borne de recharge, un switch (MitM Switch) et une entité MitM. La différence avec le scénario précédent se situe au niveau de l'entité MitM. Elle a été reconfigurée pour empêcher le SE de fournir son service. L'architecture de ce scénario se présente comme celle représentée dans la figure 4.2 ci-dessus. Les trois (03) interfaces créées par le switch sont surveillées par l'outil Wireshark lors des processus de charge du véhicule électrique. Après avoir capturé le nombre de paquets suffisants, les données sont enregistrées sous format PCAP.

Il est important de préciser que dans chaque scénario, la borne de recharge fonctionne comme un seveur et le véhicule électrique comme une machine cliente.

Suite aux différentes simulations, on se retrouve avec trois fichiers PCAP, soit un fichier PCAP par scénario. Afin d'avoir les données collectées sous format CSV et réparties suivant plusieurs variables, l'outil CICFlowMeter a été mis à contribution. CICFlowmeter est reconnu pour sa capacité à générer des datasets (base de données) relatives aux attaques. Il a été utilisé pour générer trois bases de données partielles (les fichiers PCAP enregistrés ont été importés). Chaque base de données partielles est sous format de fichier CSV avec six colonnes nommées FlowID, SourceIP, DestinationIP, SourcePort, DestinationPort, Protocol et 78 autres fonctionnalités (ou variables) de trafic réseau.

Après l'obtention de ces trois bases de données partielles, s'en suit l'opération de nettoyage desdites bases de données partielles. Pour chaque base de données partielles les variables FlowID, SourceIP, DestinationIP, SourcePort, DestinationPort et Protocol sont supprimées. Aussi, les variables (ou fonctionnalités) dont les données sont identiques pour les trois bases de données partielles ont été supprimées.

Ensuite une colonne "ATT" (attaque) a été créées dans chaque fichier CSV(bases de données partielles nettoyées).

— Dans le fichier CSV correspondant aux données relatives au scénario sans attaque : toutes les valeurs de la variable ATT sont à "0" ;
— Dans le fichier CSV correspondant aux données relatives au scénario avec une attaque de type homme du milieu : toutes les valeurs de la variable ATT sont à

"1" ;

— Dans le fichier CSV correspondant aux données relatives au scénario avec une attaque de type DoS : toutes les valeurs de la variable ATT sont à "1" ;

La variable ATT nous permet d'étiquetter nos données et de faire la différence entre les données avec et sans attaque.

Enfin après l'ajout de la variable ATT, les trois bases de données partielles nettoyées ont été fusionnées d'où l'obtention de la base de données finale.

5.2 Sélection des variables significatives

L'apport de chaque variable explicative dans la prédiction de la variable réponse ATT n'est pas le même. Certaines variables participent mieux à la prédiction de la variable binaire que d'autres. Dans le but d'éviter l'utilisation des 24 variables (obtenues suite aux opérations de nettoyage), il est important de déterminer les variables les plus significatives. La détermination des variables significatives nous permettra d'éviter la redondance des données.

La méthode StepWise sous le logiciel R a été utlisée pour déterminer les variables significatives. Mais les résultats qui en résultent ne sont pas exploitables parce que nous avons obtenu une seule variable significative qui est "Flow IAT Min".

```
             Df Deviance    AIC
<none>             0.00   4.00
- FlowIATMin  1   288.04 290.04
There were 50 or more warnings (use warnings() to see the first 50)
> # Coefficients de la régression
> coef(step.model)
(Intercept)  FlowIATMin
24.17476553 -0.01891733
```

FIGURE 5.3 – Résultats de Stepwise

Ainsi, nous avons utilisé l'ACP sous le logiciel TANAGRA afin de faire ressortir les variables significatives. Les résultats issus de l'ACP ont été concluants puisque nous avons obtenu plusieurs variables significatives comme le montre la figure 5.4 :

Attribute	Axis_1		Axis_2		Axis_3		Axis_4		Axis_5	
-	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)
d2c_Flow IAT Min_1	0,96935	94 % (94 %)	-0,01686	0 % (94 %)	0,23767	6 % (100 %)	-0,05998	0 % (100 %)	0,00000	0 % (100 %)
d2c_Bwd Pkts/s_1	0,96631	93 % (93 %)	-0,01574	0 % (93 %)	0,25526	7 % (100 %)	0,02894	0 % (100 %)	0,00000	0 % (100 %)
d2c_Fwd Pkts/s_1	0,96631	93 % (93 %)	-0,01574	0 % (93 %)	0,25526	7 % (100 %)	0,02894	0 % (100 %)	0,00000	0 % (100 %)
d2c_Flow IAT Max_1	0,95950	92 % (92 %)	0,17758	3 % (95 %)	-0,21868	5 % (100 %)	0,00065	0 % (100 %)	0,00000	0 % (100 %)
d2c_Flow IAT Mean_1	0,95950	92 % (92 %)	0,17758	3 % (95 %)	-0,21868	5 % (100 %)	0,00065	0 % (100 %)	0,00000	0 % (100 %)
d2c_Flow Pkts/s_1	0,95950	92 % (92 %)	0,17758	3 % (95 %)	-0,21868	5 % (100 %)	0,00065	0 % (100 %)	0,00000	0 % (100 %)
Flow Duration	-0,58348	34 % (34 %)	0,79591	63 % (97 %)	0,16150	3 % (100 %)	-0,00056	0 % (100 %)	0,00000	0 % (100 %)
Fwd IAT Min	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Active Min	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Idle Max	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Subflow Fwd Pkts	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
d2c_Flow IAT Std_1	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Idle Min	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Fwd IAT Max	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Active Max	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Active Std	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Idle Mean	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Fwd IAT Tot	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Fwd IAT Std	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Active Mean	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Tot Fwd Pkts	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Fwd IAT Mean	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Idle Std	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Var. Expl.	5,90954	84 % (84 %)	0,72885	10 % (95 %)	0,35634	5 % (100 %)	0,00527	0 % (100 %)	0,00000	0 % (100 %)

FIGURE 5.4 – Résultats de l'ACP

De la figure 5.4, nous voyons clairement que suite à l'ACP les axes 1 et 2 représentent les axes les plus importants car ils constituent à eux seuls 95% de la variabilité des données dont 85% (léger arrondissement) pour l'axe 1 et 10% pour l'axe 2. Aussi, les variables significatives sont : Flow IAT Min, Bwd Pkts/s, Fwd Pkts/s, Flow IAT Max, Flow IAT Mean, Flow Pkts/s qui sont liées à l'axe 1 et la variable Flow Duration qui est liée à l'axe 2.

Le nombre total de variables significatives s'élève à 7. La manipulation de ces 7 variables significatives lors de la phase d'estimation de la copule pourrait être fastidieuse pour cela, nous avons choisi trois (03) variables parmi celles qui sont liées à l'axe 1. Les variables Flow IAT Min, Bwd Pkts/s, Fwd Pkts/s sont celles retenues parce qu'elles sont les plus corrélées à l'axe 1.

Les variables retenues pour la conception de notre modèle de détection d'intrusion sont les suivantes :

— Flow IAT Min : Temps minimum entre deux flux ;

— Bwd Pkts/s : Nombre de paquets en arrière par seconde ;

— Fwd Pkts/s : Nombre de paquets de transfert par seconde ;

— Flow Duration : Durée d'écoulement ;

Dans [29], les autres fonctionnalités de CICFlowMeter sont présentées.

5.3 Méthode de prédiction

Une fois les variables significatives de notre base de données V2G obtenues, il faut maintenant concevoir un modèle capable de prédire la variable binaire "ATT". Il existe différentes techniques afin d'obtenir des modèles de prédiction d'attaques. Dans le cadre de notre travail, le modèle de prédiction d'attaques proposé est inspiré du modèle de l'article[30]. Il est basé sur la notion de régression binaire généralisée. Le but du présent modèle est de prédire en pratique une variable binaire à partir des variables explicatives.

Concrètement, soit Y une variable binaire et soit $X = (X_1, \ldots, X_n)$ un vecteur de variables explicatives

On cherche ici à estimer la probabilité prédictive d'une attaque à l'aide de l'équation suivante :

$$\pi(x) = P(Y = 1|X = x), x \in \mathbb{R}^d. \qquad (1)$$

Il s'agira à réécrire l'équation (1) en termes de distributions marginales et d'une copule. Cette réécriture est possible en utilisant le théorème de Sklar (pour plus de détail voir [30]). Pour ce faire on pose : $p = P(Y = 1)$, on a :

$$\begin{aligned} \pi(x) &= P(Y = 1|X = x) \\ &= 1 - P(Y \leq 0|X = x) \\ \pi(x) &= 1 - C^c\{(1 - p|F_1(x_1), \ldots, F_d(x_d)\}, \qquad (2) \end{aligned}$$

où $C^c\{(u|v_1, \ldots, v_d\}$ désigne la copule conditionnelle définie de manière générale par :

$$C^c\{(u|v_1,\ldots,v_d\} = \frac{\partial v_1 \ldots \partial v_d C(u, v_1, \ldots, v_d)}{\partial v_1 \ldots \partial v_d C(1, v_1, \ldots, v_d)} = \frac{1}{c(v_1, \ldots, v_d)} \int_0^y c(s, v_1, \ldots, v_d) ds,$$

L'estimation de la probabilité de prédiction décrite dans (2) peut être effectuée en cherchant séparément p, $F_1(x_1), ..., F_d(x_d)$ aussi bien que le paramètre de la copule θ. Pour ce faire, considérons $(Y_i, X_{i,1}, \ldots, X_{i,d}), i = 1, \ldots, n$, un échantillon du vecteur $(Y, X_1, \ldots, X_d)$

— Estimation de p

p est estimée par le pourcentage qu'une attaque se produise et est définie par :

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} \prod (Y_i = 1).$$

— Estimation des marginales

L'estimation des marginales est effectuée grâce à l'utilisation d'une distribution empirique redimensionnée qui est donnée par la formule :

$$\hat{F}_{j,n}(x_j) = \tfrac{1}{n+1} \textstyle\sum_{i=1}^{n} \prod (X_{i,j} \leq x_j), \quad j \in \{1, \ldots, d\}$$

Avec

$$\hat{F}_n(x) = (\hat{F}_{1,n}(x_1), \ldots, \hat{F}_{d,n}(x_d)).$$

— Estimation du paramètre de la copule θ

L'estimation est effectuée paramétriquement d'où on note θ le paramètre et $\hat{\theta}_n$ l'estimateur. On trouve l'estimateur du paramètre de notre copule en cherchant le paramètre $\hat{\theta}_n$ qui maximise la fonction de pseudo-vraisemblance suivante (en lien avec notre probabilité prédictive $\pi(x)$:

$$\mathcal{L}(\theta, p, F|Y, X) = \prod_{i=1}^{n} \ell(\theta, p, F(X_i)),$$

où

$$\ell(\theta, p, F(X_i)) = \{1 - C^c\{1 - p|F(X_i); \theta\}\}^{Y_i} C^c\{1 - p|F(X_i); \theta\}^{1-Y_i}$$

Ainsi, l'estimateur $\hat{\theta}_n$ est défini comme suit :

$$\hat{\theta}_n = \arg\max \mathcal{L}(\theta, \hat{p_n}, \hat{F}|Y, X).$$

Le package VineCopula sous R a été utilisé pour choisir la copule qui représente le mieux les données et estimer le paramètre θ le plus performant.

De tout ce qui précède l'estimateur final de $\pi(x)$ est :

$$\tilde{\pi}_n(x) = 1 - C^c\{1 - \hat{p}_n|\hat{F}_n(x); \hat{\theta}_n\}.$$

Afin de déterminer si $Y = 1$ ou $Y = 0$, il suffit par la suite de classifier de la façon

suivante : si la probabilité qu'un $Y_i = 1$ sachant les variables explicatives (x) est plus grande que 50% et que c'est soit $Y = 0$ ou $Y = 1$, alors il faut associer Y_i à 1.

Dans le cadre de la réalisation du modèle sous le logiciel R, la méthodologie suivante inspirée de [31] équation (6) a été utilisée pour estimer le paramètre de la copule.

En appliquant la formule (6) de [31] et en considérant le fait que nous sommes en présence de quatre variables significatives, cette équation peut être réécrite de la manière suivante :

$$F(Y \mid X_1, X_2, X_3, X_4) = \frac{\partial C(F(Y \mid X_2, X_3, X_4), F(X_1 \mid X_2, X_3, X_4))}{\partial F(X_1 \mid X_2, X_3, X_4)} ;$$

Dans cette expression, $F(Y \mid X_2, X_3, X_4)$ et $F(X_1 \mid X_2, X_3, X_4)$ sont des inconnues qu'il faudra chercher. Pour le faire, il suffit de suivre les étapes suivantes :

1ère étape : Calculer $F(X_2 \mid X_4)$, $F(X_3 \mid X_4)$, $F(X_1 \mid X_4)$ et $F(Y \mid X_4)$

$$F(X_2 \mid X_4) = \frac{\partial C(F(X_2), F(X_4))}{\partial F(X_4)} \quad ; \quad F(X_1 \mid X_4) = \frac{\partial C(F(X_1), F(X_4))}{\partial F(X_4)}$$

$$F(X_3 \mid X_4) = \frac{\partial C(F(X_3), F(X_4))}{\partial F(X_4)} \quad ; \quad F(Y \mid X_4) = \frac{\partial C(F(Y), F(X_4))}{\partial F(X_4)}$$

2ème étape : Utiliser les expressions de $F(X_2 \mid X_4)$, $F(X_3 \mid X_4)$, $F(X_1 \mid X_4)$ et $F(Y \mid X_4)$ pour calculer $F(X_2 \mid X_3, X_4)$, $F(X_1 \mid X_3, X_4)$ et $F(Y \mid X_3, X_4)$.

$$F(X_2 \mid X_3, X_4) = \frac{\partial C(F(X_2 \mid X_4), F(X_3 \mid X_4))}{\partial F(X_3 \mid X_4)} ;$$

$$F(X_1 \mid X_3, X_4) = \frac{\partial C(F(X_1 \mid X_4), F(X_3 \mid X_4))}{\partial F(X_3 \mid X_4)} ;$$

$$F(Y \mid X_3, X_4) = \frac{\partial C(F(Y \mid X_4), F(X_3 \mid X_4))}{\partial F(X_3 \mid X_4)} ;$$

3ème étape : Ensuite, utiliser les expressions calculées à l'étape 2 pour déterminer $F(X_1 \mid X_2, X_3, X_4)$ et $F(Y \mid X_2, X_3, X_4)$.

$$F(X_1 \mid X_2, X_3, X_4) = \frac{\partial C(F(X_1 \mid X_3, X_4), F(X_2 \mid X_3, X_4))}{\partial F(X_2 \mid X_3, X_4)} ;$$

$$F(Y \mid X_2, X_3, X_4) = \frac{\partial C(F(Y \mid X_3, X_4), F(X_2 \mid X_3, X_4))}{\partial F(X_2 \mid X_3, X_4)} ;$$

4ème étape : Les expressions $F(X_1 \mid X_2, X_3, X_4)$ et $F(Y \mid X_2, X_3, X_4)$ vont nous aider à déterminer $F(Y \mid X_1, X_2, X_3, X_4)$.

5.4 Conclusion

En conclusion, la méthode de prédiction proposée s'applique à des données à n-dimensions. La première étape de cette méthode consiste à estimer la probabilité qu'il y ait une attaque sous la forme d'une équation. Ainsi, l'équation obtenue est basée sur la probabilité p, les marginales et la copule. Ensuite, une estimation de chacun de ces paramètres est déterminée. Enfin, pour savoir s'il s'agit d'une attaque ou non, une classification est faite. Dans le cadre de nos travaux, cette méthode a été appliquée à une base de données issue de la collecte d'informations dans le réseau V2G simulé.

Chapitre 6

ARTICLE SCIENTIFIQUE

A COPULA-BASED ATTACK PREDICTION MODEL FOR V2G NETWORKS

Soumis au journal Applied Sciences de l'éditeur MDPI.

Publié le 7 avril 2022.

Référence : Nonvignon, T.Z. ; Boucif, A.B. ; Mhamed, M. ; A Copula-Based Attack Prediction Model for Vehicle-to-Grid Networks ; Appl. Sci. 2022, 12, 3830.

applied sciences

MDPI

Article

A Copula-Based Attack Prediction Model for Vehicle-to-Grid Networks

Toffa Zidane Nonvignon *, Amar Bensaber Boucif and Mesfioui Mhamed

Departement of Mathematics and Computer Science, Université du Québec à Trois-Rivières, Trois-Rivières, QC G8Z 4M3, Canada; boucif.amar.bensaber@uqtr.ca (A.B.B.); mhamed.mesfioui@uqtr.ca (M.M.)
* Correspondence: zidane.nonvignon.toffa@uqtr.ca

Abstract: The Vehicle-to-Grid (V2G) networks are a part of the Smart Grid networks. Their primary goal is to recharge electric vehicles. These networks, as with any computer system, are facing cyber attacks. For example, during a charge or recharge process, V2G networks can be vulnerable to attacks such as Man-in-the-Middle (MitM), Denial of Service (DoS), identity theft, and rebound attacks. It is therefore up to us to offer innovative solutions in order to reduce threats as much as possible. In this paper, a model based on copulas to detect intrusion cases in V2G networks is proposed. To achieve this model, a database is generated first from three scenarios using tools including MiniV2G, Wireshark, and CICflowMeter. Then, significant variables are selected using Principal Component Analysis (PCA). The classification algorithm is based on the notion of copulas constructed under the software R. From the obtained results, it emerges that the created model has a very high prediction rate of attacks in the aforementioned network.

Keywords: V2G security; attack prediction; dataset; copula

Citation: Nonvignon, T.Z.; Boucif, A.B.; Mhamed, M. A Copula-Based Attack Prediction Model for Vehicle-to-Grid Networks. *Appl. Sci.* **2022**, *12*, 3830. https://doi.org/10.3390/app12083830

Academic Editor: Fabrizio Granelli

Received: 12 March 2022
Accepted: 7 April 2022
Published: 11 April 2022

1. Introduction

Global warming is a phenomenon manifested by an increase in average temperatures due to significant greenhouse gas emissions. It constitutes a situation of a global nature and is worrying such that all the nations of the world are mobilized to find the ideal solutions to preserve the environment and leave a livable planet for future generations. Gasoline vehicles are one of the main sources of greenhouse gas emissions. It becomes urgent to move towards electric vehicles and renewable energy sources. Faced with these challenges, the Smart Grid, or the intelligent electricity grid, has emerged. The Smart Grid was born from the merger between the conventional electricity network and information systems. It promotes renewable energies and manages the balance between production and consumption. However, as with any system, the Smart Grid—and, more precisely, the V2G—has vulnerabilities that can be exploited by malicious parties. The V2G is a part of the Smart Grid; it is responsible for bidirectional exchanges between the electric vehicle and its charging station.

As with other networks, V2G networks need solutions in order to secure communications between the vehicle and the charging station. Thus, one obstacle to the large-scale deployment of electric vehicles would be overcome. To resolve the security issue in V2G, it is possible to draw inspiration from Intrusion Detection Systems (IDS). These IDS have proven their worth in the networks that preceded the V2G networks.

The objective of our work is to propose a model capable of predicting DoS and MitM attacks. This model is based on copulas.

The rest of this paper is organized as follows. In Section 2, general information about V2G networks is exposed. In Section 3, the state of the art on security requirements in V2G networks and methods to detect attacks in multiple networks are presented. In Section 4, the model is described. In Section 5, the process of obtaining the attack database is presented, and finally, the obtained results are discussed in Section 6.

2. General Information on V2G Networks

A V2G network is one that allows electric vehicles (battery electric vehicles, plug-in hybrid vehicles) to recharge or distribute a portion of their electrical energy to the electricity grid. The electrical energy that passes through the V2G comes from renewable energy sources such as solar and wind. The use of this technology promotes the development of another means to make money and cars become less polluting (emitting less CO_2).

According to the 15118-1 standard, the entities of the V2G vehicular network are classified into two major categories. The first category takes into account the primary actors and the other category the secondary actors.

The primary actors are directly involved in all the services provided by the aforementioned network and are listed as follows:

- the electric vehicle, which is composed of a battery, a communication controller, a man–machine interface, and an electronic control unit;
- the charging station, which consists of a communication controller, an electric meter, a man–machine interface, and a payment unit.

It should be noted that each actor mentioned above has a communication controller. The communication controller is essential to establish, maintain, and terminate communication between them.

Concerning the secondary actors, they intervene indirectly in the exchange process. The secondary actors are the mobility operator, the electricity supplier, and the car manufacturer.

As with any computer system, the V2G network is confronted by several types of attacks. These attacks include:

- DoS (consists of making a hardware or software resource unavailable to clients by overloading the network with requests);
- MitM (consists of an outside entity intercepting communications between two entities);
- Identity theft (involves falsifying communication between an electric vehicle and its charging station in order to impersonate one or the other);
- The rebound attack (involves attacking an electric car or charging station through another machine in order to hide its tracks).

To protect a V2G from these attacks, IDS and secure architectures (such as Public Key Infrastructure) have been proposed by various researchers [1,2].

The role of IDS is to detect suspicious or abnormal activities on the basis of data collected in a network (here, it is the V2G network). These systems are generally classified into two categories.

- Signature-based IDSs:
  For this category, it is important to have an up-to-date attack signature database to avoid attacks. However, a signature-based IDS does not protect against unknown or unlisted attacks.
- IDSs based on abnormalities:
  This is the ability of the IDS to differentiate abnormal behavior from normal behavior. In order to have this capacity, the IDS would have to be trained beforehand. To carry out the training of an IDS, several techniques are used. These involve machine learning [3] and deep learning [4].

However, it is possible to encounter a hybrid IDS, which represents a combination of the two types. In this work, an attack prediction model that relates to an IDS based on abnormalities is proposed.

3. State of the Art

The United States plans to invest approximately $28 billion between 2021 and 2030 to reach 2.4 million charging stations by 2030 [5]. Moreover, Canada, through its Zero Emission Vehicle Infrastructure Program (ZEVIP), has decided to invest, between 2019 and

2024, $280 million to remedy the lack of charging stations [6]. It is therefore important to know the security requirements of the V2G network and to have reliable solutions in order to minimize the risk of network compromise by cybercriminals during the large-scale deployment phase.

In this section, a state of the art is presented on the security objectives of the V2G network and some solutions proposed in the literature to detect and thwart attacks on several networks.

3.1. Security Challenges in V2G Networks

The issue relating to the security aspect in vehicular networks and, more specifically, in the V2G network is a topical and worrying issue. For this purpose, the International Telecommunications Union (ITU), in accordance with its prerogatives, has issued guidelines relating to the security of vehicle communications to any other element (V2X) through its recommendation [7]. In this recommendation, it mainly concerns the threats encountered in the operation of V2X communications systems and the security requirements for this type of communication. In addition, it provides guidance in implementing secure V2X communications.

The vulnerabilities of a network are exploited by hackers to carry out attacks in order to obtain unauthorized access to a system, steal personal or confidential information, disrupt the proper functioning of a service, collect data, and many more. Based on the recommendation cited above, a V2X network (which takes into account the V2G network) is exposed to many attacks: MitM, DoS, spoofing, and time attacks. Moreover, we can deduce from this recommendation that confidentiality, integrity, availability, non-repudiation, authenticity, and accountability are security requirements that must be implemented in all proposed solutions for V2G.

In [8], the authors discussed concepts such as basic concepts, architecture, advantages, and cybersecurity in V2G networks. A model based on cyber-insurance has been proposed. The goal of this solution is to transfer the risk of compromise of an electric vehicle to an insurer, who will be responsible for putting in place the necessary controls to protect the asset against cyber attacks and to respond to incidents.

In order to achieve cybersecurity objectives such as confidentiality, authenticity, and integrity in the V2G network, an architecture has been proposed in [9]. This proposed architecture is composed of electric vehicles, charging stations, aggregators, communication servers, and authentication servers. With features such as anonymous authentication, anonymous signatures, and remote attestation, the proposed architecture preserves the privacy of EVs and secures charging and discharging processes.

In the field of V2G security, the International Organization for Standardization (ISO) 15118 standard is the reference standard. It deals with issues related to secure communications between an electric vehicle (EV) and a charging station. This standard is divided across several versions:

- ISO 15118-1: discusses the generalities and use cases of the ISO 15118 standard [10];
- ISO 15118-2: discusses the requirements for Layer 3 to 7 protocols of the OSI model [11];
- ISO 15118-3: discusses the requirements for Layer 1 and Layer 2 protocols of the OSI model [12].

However, it should be noted that the ISO 15118 standard, although it is a reference in terms of security in V2G, does not take into account all aspects relating to security. It does not give a precise orientation on all the solutions and methods to be implemented in order to protect V2G. The limits of the latter have been corrected by researchers, who propose several solutions to strengthen security in V2G. IDSs are considered one of the solutions to protect the V2G networks.

3.2. Detection of Attacks in Networks

In [13], the authors investigated the various modern techniques used to detect an intrusion in a network. Modern techniques used in the detection of anomalies in networks

are machine learning, deep learning, and blockchain technology. They have a higher detection rate. To shed light on any malicious activity, machine learning relies on algorithms such as Support Vector Machine (SVM), Bayes classifier, decision table, decision tree, clustering, and k-means, while learning in depth relies on algorithms such as Auto Encoder, Convolutional Neural Network (CNN), Deep Neural Network (DNN), and Boltzmann Machine (BM).

In [14], Petros Toupas et al. suggested an intrusion detection model for the Internet network that is based on a deep neural network. As part of the realization of the intrusion detection model, downloading the CICIDS2017 database was the first step. Then, cleaning and standardization operations were carried out on the CICIDS2017 database. The new database (cleaned CICIDS2017) obtained after the cleaning operations has 44 variables, which constitute input data for a deep neural network. The entry layer is followed by eight layers composed, respectively, of 140, 120, 100, 80, 60, 40, 20, and 120 nodes. The final layer is the output layer that produces the probabilities. The ability of the model to predict attacks was assessed using measurement indicators such as precision, recall, and F-measure. The results of this evaluation prove that the attack detection model is effective. For example, accuracy is 99.95%, the precision is equal to 94.31%, the recall is 95.62%, and F-measure is 94.1%.

In [15], Caroly Gabriela Pereira Diaz et al. proposed an adaptive neuro-fuzzy inference system (ANFIS) to obtain a hybrid model for predicting a safety indicator in Vehicular Ad-Hoc Networks (VANET). This model consists of using a neural network that is built on five layers coupled with a fuzzy inference system of the Takagi Sugeno type. For the realization of such a model, the first step was the production of a database. This database is obtained after collecting information from simulations of two scenarios. In the first scenario, the simulated network works without an attack, and the network is subjected to a DoS attack. Then, a correlation test based on the Pearson method is performed on the selected variables, followed by a PCA in order to identify the significant variables. Finally, the ANFIS is defined and built using MATLAB Toolbox. The results of this work show that the ANFIS model works quite well in the attack detection process.

To detect and prevent attacks in the VANET network, the authors in [16] have proposed a model based on the notion of Watchdog and the theory of Bayesian networks. The Watchdog method is used to detect malicious nodes. Indeed, nodes with a watchdog are responsible for constantly listening to neighboring nodes and monitoring their behavior. With this method, it was possible to end up with a high rate of false positives due to nodes' mobility and signal noise. To reduce the high rate of false positives, a Bayesian filter was used by the authors.

To combat DoS-type attacks in VANET networks, Deepak Rampaul et al. studied several methods [17]. These are methods such as detecting jamming attacks in vehicle ad hoc network (DJVAN), a signature-based authentication method, and the Request Response Detection Algorithm (PRDA). From this work, it appears that, despite the techniques for fighting against DoS attacks, it is currently not possible to make DoS attacks disappear.

In [18], the authors proposed an IDS based on machine learning to fight against fuzzy and DoS attacks. To perform the IDS, the authors relied on two types of data, "fuzzy attack data" and "DoS data", which were provided by the Hacking and Countermesasure Research Lab (HCRL). On these data, cleaning and normalization operations were performed. Then, the data were labeled to differentiate the normal data from the injected data. After this step, the SVM and K-Nearest Neighbors (KNN) algorithms were implemented for classification. The model, once trained, is able to detect DoS and fuzzy attacks. As part of the assessment, a comparison was made to see which classification algorithm worked best. It emerged from this comparison that the two algorithms gave excellent results; however, the KNN algorithm is more efficient than the SVM algorithm.

In [19], a mechanism is proposed to fight against Sybil-type attacks in a VANET network. The Sybil attack involves generating fake vehicles around a legitimate vehicle. The proposed solution is based on an existing technology called the "Advanced Driving

Assistant System (ADAS) sensor", which allows a car that has it to control its environment. The proposed detection mechanism takes place in three consecutive stages. The first step is to validate the messages, i.e., to ensure that the received message comes from a valid source. Then, the second step is to verify the sender and its location. The last step is to check the surrounding objects. From the analysis of the obtained results from the CARLA simulator, it emerges that a legitimate vehicle is able to detect false vehicles on a reduced distance. Over a long distance, there is a high number of false positives.

Beyond the various methods mentioned above on which IDSs are based, there are the mathematical methods. Attack detection systems can be based on mathematical models. In [20], the authors proposed a model for the detection of anomalies based on copulas. This model is based firstly on the determination of the maximum information tree; then, the modeling of the joint distributions of the pairs of variables obtained by the maximum information tree is performed, and finally the attribution of the anomaly scores. The evaluation of the model, carried out using a database from the base stations of an LTE network in Europe, gave very satisfactory results. Moreover, in the same work, the copula-based anomaly detection model gave better results compared to other methods such as a PCA-based anomaly detection method, isolation forest, the anomaly detection method based on the local outlier factor, the anomaly detection method based on clustering-LOF, and the anomaly detection method based on the auto-encoder.

In [21], the authors have proposed mathematical methods in order to detect DoS-type attacks in the VANET network. They used and compared "Root Mean Square (RMS)", "Mean Absolute Value (MAV)", "Mean Squared Error (MSE)", and "Logistic Regression Model". They also proposed a neural network model to protect the entities of the VANET network against DoS attacks. These different methods were applied to a database resulting from a collection of information relating to simulations carried out using the OMNET ++ and SUMO tools. Before using the database, cleaning operations were carried out. After analysis of the various obtained results, it emerged that the logistic regression and the neural network are better than the MSE, RMS, and MAV in the context of the prediction of DoS-type attacks in a VANET network.

From the above, for the detection of attacks in networks, several approaches have been used. These include models based on neural networks, deep learning, or mathematics. In the next section, the method of attack prediction in V2G networks is presented. It is based on copulas.

4. Prediction Method

In the context of our work, the attack prediction method proposed is inspired by the model presented in [22]. It is based on the notion of generalized binary regression. The aim of this model is to predict, in practice, a binary response variable through the explanatory variables. Specifically, let Y be a binary variable taking values 0 and 1, and let $X = (X_1, \ldots, X_d)$ be a vector of explanatory variables. The goal is to estimate the predicted probability of an attack using the following equation:

$$\pi(x) = \mathrm{P}(Y = 1 | X = x), \quad x \in \mathbb{R}^d. \tag{1}$$

In the following, a new way to model the above predicted probability using the notion of copulas is presented. This idea consists of rewriting $\pi(x)$ in terms of the marginal distribution and the copula of the vector $(Y, X_1, \ldots, X_d)$. This can be done using the Sklar Theorem (see [22] for more details). In fact, let $p = \mathrm{P}(Y = 1)$; then, one observes that

$$\begin{aligned} \pi(x) &= P(Y = 1 | X = x) \\ &= 1 - P(Y \leq 0 | X = x) \\ &= 1 - C^c\{1 - p | F_1(x_1), \ldots, F_d(x_d)\}, \end{aligned} \tag{2}$$

where $C^c\{u|v_1,\ldots,v_d\}$ denote the conditional copula expressed by

$$C^c\{u|v_1,\ldots,v_d\} = \frac{c(u,v_1,\ldots,v_d)}{c(1,v_1,\ldots,v_d)},$$

with

$$c(u,v_1,\ldots,v_d) = \frac{\partial^d C}{\partial v_1 \cdots \partial v_d}(u,v_1,\ldots,v_d).$$

The estimation of the predicted probability described in (2) can be achieved by estimating separately the marginal quantities, $p, F_1(x_1),\ldots,F_d(x_d)$, as well as the copula parameter θ. This means that the estimation procedure can be executed through the next steps. For this, let $(Y_i, X_{i,1},\ldots,X_{i,d})$, $i = 1,\ldots,n$ be a sample from the random vector $(Y, X_1,\ldots,X_d)$.

- Estimation of p.
  The probability p can simply be estimated by the percentage of an attack, given by

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n}\prod(Y_i = 1).$$

- Estimation of marginal.
  The marginal distributions $F_1(x_1),\ldots,F_d(x_d)$ are estimated using empirical distribution:

$$\hat{F}_{j,n}(x_j) = \frac{1}{n+1}\sum_{i=1}^{n}\mathbb{I}(X_{i,j} \le x_j), \quad j = 1,\ldots,d.$$

- Estimation of the copula parameter θ.
  The estimator $\hat{\theta}_n$ of the dependence parameter θ is derived from the maximum likelihood procedure. In fact, $\hat{\theta}_n$ is obtained by maximizing in terms of θ the following pseudo-likelihood function:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n}\ell(\theta,\hat{p},Y_i,F_n(X_i)),$$

  where $F_n(X_i) = (F_{1,n}(X_{i,1}),\ldots,F_{d,n}(X_{i,d}))$ and $\ell(\theta,\hat{p},Y_i,F_n(X_i))$ denotes

$$\{1 - C^c\{1-\hat{p}|F_n(X_i);\theta\}\}^{Y_i}C^c\{1-\hat{p}|F_n(X_i);\theta\}^{1-Y_i}.$$

  Consequently, the desired estimator $\hat{\theta}_n$ is given by

$$\hat{\theta}_n = \arg\max \mathcal{L}(\theta).$$

The VineCopula package under software R was used to choose the copula that best represents the data and estimate the best-performing θ parameter.

From all of the above, the final estimator of $\pi(x)$ is

$$\tilde{\pi}_n(x) = 1 - C^c\{1-\hat{p}_n|F_n(x);\hat{\theta}_n\}$$

In order to determine whether $Y = 1$ or $Y = 0$, it suffices thereafter to classify as follows: if the probability that $Y_i = 1$ knowing the explanatory variables x is greater than 50% and it is either $Y = 0$ or $Y = 1$, then we must associate Y_i with 1.

Within the framework of the realization of the model under the software R, the following methodology inspired by [23] Equation (6) was used to estimate the parameter of the copula.

By applying Formula (6) of [23] and considering four significant variables X_1, X_2, X_3, X_4, the distribution of (Y, X_2, X_3, X_4) can be rewritten as follows:

$$F(Y|X_1,X_2,X_3,X_4) = \frac{\partial C(F(Y|X_2,X_3,X_4),F(X_1|X_2,X_3,X_4))}{\partial F(X_1|X_2,X_3,X_4)}.$$

In this expression, the conditional distributions $F(Y|X_2, X_3, X_4)$ and $F(X_1|X_2, X_3, X_4)$ are unknowns that must be sought. To do so, we simply perform the following steps:

1st step: Compute the conditional distributions $F(X_2|X_4)$, $F(X_3|X_4)$, $F(X_1|X_4)$ and $F(Y|X_4)$

$$F(X_2|X_4) = \frac{\partial C(F(X_2), F(X_4))}{\partial F(X_4)} \ ;$$

$$F(X_1|X_4) = \frac{\partial C(F(X_1), F(X_4))}{\partial F(X_4)} \ ;$$

$$F(X_3|X_4) = \frac{\partial C(F(X_3), F(X_4))}{\partial F(X_4)} \ ;$$

$$F(Y|X_4) = \frac{\partial C(F(Y), F(X_4))}{\partial F(X_4)} \ .$$

2nd step: Use the expressions of $F(X_2|X_4)$, $F(X_3|X_4)$, $F(X_1|X_4)$ and $F(Y|X_4)$ to calculate $F(X_2|X_3, X_4)$, $F(X_1|X_3, X_4)$ and $F(Y|X_3, X_4)$.

$$F(X_2|X_3, X_4) = \frac{\partial C(F(X_2|X_4), F(X_3|X_4))}{\partial F(X_3|X_4)};$$

$$F(X_1|X_3, X_4) = \frac{\partial C(F(X_1|X_4), F(X_3|X_4))}{\partial F(X_3|X_4)};$$

$$F(Y|X_3, X_4) = \frac{\partial C(F(Y|X_4), F(X_3|X_4))}{\partial F(X_3|X_4)}.$$

3rd step: Then, use the expressions calculated in step 2 to determine $F(X_1|X_2, X_3, X_4)$ and $F(Y|X_2, X_3, X_4)$.

$$F(X_1|X_2, X_3, X_4) = \frac{\partial C(F(X_1|X_3, X_4), F(X_2 \mid X_3, X_4))}{\partial F(X_2|X_3, X_4)};$$

$$F(Y|X_2, X_3, X_4) = \frac{\partial C(F(Y|X_3, X_4), F(X_2 \mid X_3, X_4))}{\partial F(X_2|X_3, X_4)}.$$

4th step: The expressions $F(X_1|X_2, X_3, X_4)$ and $F(Y|X_2, X_3, X_4)$ will help us to determine $F(Y|X_1, X_2, X_3, X_4)$.

In conclusion, the prediction method is applicable to n-dimensional data and can be summarized as follows: estimate the probability that there is an attack in the form of an equation. Thus, the equation obtained is based on the probability p, the marginals, and the copula. Then, an estimate of each of these parameters is determined. Finally, to determine where there is an attack or not, a classification is made. As part of our work, this method was applied to a database resulting from the collection of information in the simulated V2G network.

5. V2G Network Simulation

In this section, the process for obtaining the attack database and the approach to select the significant variables are presented. The use of the MiniV2G emulator, Wireshark, and CICFlowMeter allowed us to generate an attack database. This database underwent cleaning operations and then, by using PCA, the significant variables were selected.

5.1. Generation of the Attack Database

To be able to detect attacks in a network and, more precisely, in the V2G network, it is essential to have an attack database resulting from the collection of information exchanged between network entities. In the absence of this database, it must be generated following several simulations. Thus, the open-source tool called "MiniV2G", proposed by Luca Attanasio et al. [24] and built on Mininet and RiseV2G, allowed us to simulate a V2G environment, taking into account the requirements of the ISO 15118 standard.

The simulations were carried out on the basis of three different scenarios.

(a) Scenario without attack (Figure 1).
In this scenario, the entities used are: a charging station, a switch (s1), a controller (c0), and two electric vehicles.

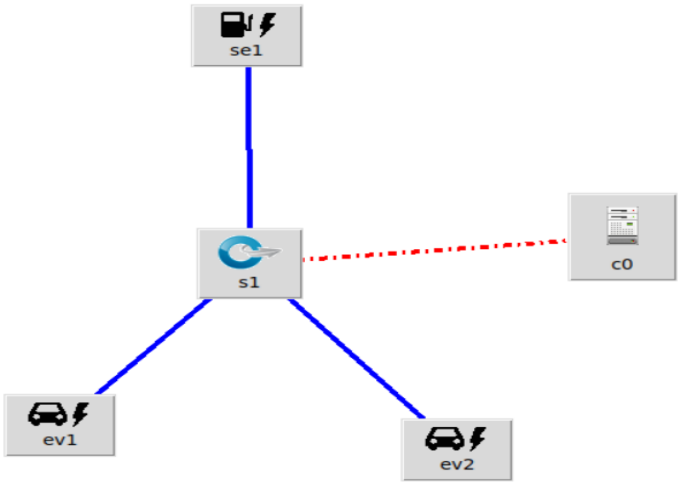

Figure 1. Attack-free simulation architecture.

(b) Scenario with a MitM attack (Figure 2).
The purpose of the MitM Switch is to intercept traffic leaving the Electric Vehicle Communication Controller (EVCC) and redirect it to the MitM node. The MitM node is able to perform any kind of manipulation before sending the network flow to the Supply Equipment Communication Controller (SECC).

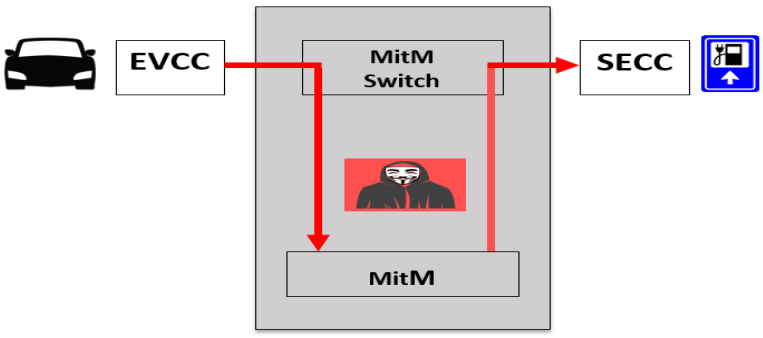

Figure 2. Architecture relating to the simulation of the MitM-type attack.

(c) Scenario with a DoS-type attack.
The architecture of this scenario resembles the one shown in the previous figure above. The MitM entity has been reconfigured and its function is to impede the charging process of the electric vehicle.

In each scenario, the interfaces created by each entity in the network are monitored by the Wireshark tool during the process of charging electric vehicles. The charging process was repeated several times in order to collect enough packets. After sorting the interfaces, the data are saved in packet capture (PCAP) format. At the end of these three simulations, three PCAP files are obtained. In order to have the data collected in Comma-Separated Values (CSV) format and distributed according to several variables, the CICFlowMeter tool was used.

CICFlowMeter is recognized for its ability to generate datasets (database) relating to attacks. It has been used to generate three partial databases (recorded PCAP files have been imported). Each partial database is in CSV file format with columns labeled for each flow, namely: FlowID, SourceIP, DestinationIP, SourcePort, DestinationPort, Protocol, and 78 other network traffic features (or variables).

After obtaining three partial databases that correspond to the three simulations carried out, the operation of cleaning these databases is the next step. For each partial database, the variables FlowID, SourceIP, DestinationIP, SourcePort, DestinationPort, Protocol, and the variables (or functionalities) whose data are identical for the three CSV files have been deleted. Then, the data in the database were labeled with the variable ATT (abbreviation of the word attack), added in order to differentiate between the data without attack and with attack. Thus, we have the following:

- In the CSV file corresponding to the data relating to the scenario without attack variable, ATT = 0;
- In the CSV file corresponding to the data relating to the scenario with a MitM attack, ATT = 1;
- In the CSV file corresponding to the data relating to the scenario with a DoS type attack, ATT = 1.

Finally, after adding the ATT variable, the three partial databases were merged, from which the final database was obtained.

5.2. Selection of Significant Variables

The contribution of each explanatory variable in the prediction of the binary variable ATT is not the same. Some variables participate better in the prediction of the ATT response variable than others. In order to avoid the manipulation of the 24 variables (obtained after the cleaning operation) during the production phase of our model, it was important to determine the significant variables. Determining the significant variables allowed us to avoid data redundancy.

The StepWise method under R software was used to determine the significant variables. However, the results were not exploitable because only one significant variable was found. Thus, the PCA method under the TANAGRA software was used in order to bring out the significant variables. The results of the PCA were conclusive since several significant variables were found, as shown in Figure 3 below.

Attribute	Axis_1		Axis_2		Axis_3		Axis_4		Axis_5	
-	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)
d2c_Flow IAT Min_1	0,96935	94 % (94 %)	-0,01686	0 % (94 %)	0,23767	6 % (100 %)	-0,05998	0 % (100 %)	0,00000	0 % (100 %)
d2c_Bwd Pkts/s_1	0,96631	93 % (93 %)	-0,01574	0 % (93 %)	0,25526	7 % (100 %)	0,02894	0 % (100 %)	0,00000	0 % (100 %)
d2c_Fwd Pkts/s_1	0,96631	93 % (93 %)	-0,01574	0 % (93 %)	0,25526	7 % (100 %)	0,02894	0 % (100 %)	0,00000	0 % (100 %)
d2c_Flow IAT Max_1	0,95950	92 % (92 %)	0,17758	3 % (95 %)	-0,21868	5 % (100 %)	0,00065	0 % (100 %)	0,00000	0 % (100 %)
d2c_Flow IAT Mean_1	0,95950	92 % (92 %)	0,17758	3 % (95 %)	-0,21868	5 % (100 %)	0,00065	0 % (100 %)	0,00000	0 % (100 %)
d2c_Flow Pkts/s_1	0,95950	92 % (92 %)	0,17758	3 % (95 %)	-0,21868	5 % (100 %)	0,00065	0 % (100 %)	0,00000	0 % (100 %)
Flow Duration	-0,58348	34 % (34 %)	0,79591	63 % (97 %)	0,16150	3 % (100 %)	-0,00056	0 % (100 %)	0,00000	0 % (100 %)
Fwd IAT Min	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Active Min	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Idle Max	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Subflow Fwd Pkts	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
d2c_Flow IAT Std_1	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Idle Min	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Fwd IAT Max	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Active Max	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Active Std	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Idle Mean	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Fwd IAT Tot	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Fwd IAT Std	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Active Mean	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Tot Fwd Pkts	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Fwd IAT Mean	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Idle Std	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)	0,00000	0 % (0 %)
Var. Expl.	5,90954	84 % (84 %)	0,72885	10 % (95 %)	0,35634	5 % (100 %)	0,00527	0 % (100 %)	0,00000	0 % (100 %)

Figure 3. PCA results.

From this figure, axis 1 and axis 2 represent the most important axes because they alone constitute 95% of the variability of the data, namely 85% (slight rounding) for axis 1 and 10% for axis 2. Moreover, the significant variables are Flow IAT Min, Bwd Pkts/s, Fwd Pkts/s, Flow IAT Max, Flow IAT Mean, Flow Pkts/s, which are related to axis 1, and the Flow Duration variable, which is linked to axis 2.

The total number of significant variables is 07. The manipulation of these 07 significant variables during the estimation phase of the copula of our model could be tedious. For this, three variables were chosen, among those that were related to axis 1, namely the variables Flow IAT Min, Bwd Pkts/s, Fwd Pkts/s, were those used because they were the most correlated with axis 1.

Table 1 presents an overview of the features of CICFlowMeter. More features of CICFlowMeter are presented in [25].

Table 1. Some features of CICFlowMeter.

Variables	Description
Bwd Pkts/s	Number of backward packets per second
Flow Duration	Flow time
Flow IAT Max	Maximum time between two flows
Flow IAT Mean	Average time between two flows
Flow IAT Min	Minimum time between two flows
Flow Pkts/s	Flow packets rate that is number of packets transferred per second
Fwd Pkts/s	Number of forward packets per second

Summary

The use of tools such as MiniV2G, Wireshark, and CICFlowMeter led to obtaining the V2G attack database. MiniV2G was used to simulate three types of scenarios (without attack, MitM type attack, DoS-type attack); Wireshark was used to assess the network in each scenario, and CICFlowMeter relied on Wireshark's PCAP files to generate the databases, which were then cleaned. PCA facilitated the detection of significant variables.

The following section presents the obtained results.

6. Results and Discussion

In this section, the results obtained following the prediction of the binary variable "ATT" (abbreviated attack) with the intrusion detection model based on the copulas are presented. The explanatory variables of our model are the four significant variables determined above.

It is important to specify that our model was developed using R software and that the information contained in our generated V2G database is divided into two parts. The first part contains 80% of the data for the training and the other part is made up of 20% of the test data.

The obtained results are presented below:

1. a prediction rate of 96.43%, which is equivalent to an error rate of 3.57%;
2. the confusion matrix of our model, which is shown in the Table 2.

Table 2. Confusion matrix for the V2G network. TN: True Negative; FN: False Negative; FP: False Positive; TP: True Positive.

		Actual values		
		0	1	**Total**
Predicted values	0	38 (TN)	0 (FN)	38
	1	2 (FP)	16 (TP)	18
	Total	40	16	56

3. $\text{Recall} = \dfrac{\text{TP}}{\text{TP}+\text{FN}} = \dfrac{16}{16+0} = 1$

   This result means that, at the level of our sample of test data, our model was able to predict all of the data relating to attacks.

4. $\text{Precision} = \dfrac{\text{TP}}{\text{TP}+\text{FP}} = \dfrac{16}{16+2} = 0.88$

   It can be said that 88% of the data that are predicted as attacks are actually attacks.

5. F-measure is considered to be a combination of recall and precision. It is then optional in our case:

$$\text{F-measure} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} = 2 \times \frac{1 \times 0.88}{1 + 0.88} = 0.93$$

From these different obtained results, the model of intrusion detection based on copulas gave very satisfactory results.

These results were confirmed during the evaluation of our model with another database generated by the authors in [15]. It is an attack database generated from data collected in the VANET vehicular network.

The obtained results after the evaluation of our model with the VANET database are presented as follows:

1. a prediction rate of 95.83%, which is equivalent to an error rate of 4.17%;

2. the confusion matrix of our model, which is shown in the Table 3.

Table 3. Confusion matrix relating to the VANET network. TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative.

		Actual values		
		0	1	**Total**
Predicted values	0	38 (TN)	0 (FN)	38
	1	2 (FP)	8 (TP)	10
	Total	40	8	48

3. $\text{Recall} = \dfrac{\text{TP}}{\text{TP} + \text{FN}} = \dfrac{8}{8+0} = 1$

4. $\text{Precision} = \dfrac{\text{TP}}{\text{TP} + \text{FP}} = \dfrac{8}{8+2} = 0.8$

5. $\text{F-measure} = 2 \times \dfrac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} = 2 \times \dfrac{1 \times 0.8}{1 + 0.8} = 0.89$

From the analysis of the various obtained results, the model based on copulas is able to predict attacks better in a V2G network than in a VANET network, which is quite normal since cars in VANET networks are moving at different speeds.

7. Conclusions

The decentralization of electric vehicles in the coming years will require the multiplication of recharging infrastructures. It would then be essential to detect attacks carried out by malicious parties in the V2G network. To meet this challenge, it is important to offer innovative intrusion detection solutions capable of meeting security requirements such as confidentiality, integrity, availability, and non-repudiation. Once the security barrier has been lifted, the population's support for electric vehicles will be high and so a major step will be taken by humans in their environmental protection program.

In this paper, an attack prediction model for the V2G network based on the notion of copulas is proposed. This model is inspired by the work of Mhamed Mesfioui et al. [22]. The different phases of realization of our model are presented as two main stages, namely the generation of the V2G attack database following the various simulations and the design of the algorithm responsible for the prediction.

In this work, DoS- and MitM-type attacks are highlighted. In addition, we evaluated whether DoS or MitM was predicted, with a prediction rate of 96.43% using the prediction method based on copulas. Our model protects V2G networks from DoS and MitM attacks only; it is not able to fight against identity theft and rebound attacks. Continuing from the present study, it would be interesting to find the source of an attack in the network in order to reduce the response time following an incident.

Author Contributions: Conceptualization, M.M. and A.B.B.; methodology, A.B.B. and T.Z.N.; software, T.Z.N.; validation, M.M. and A.B.B.; formal analysis, T.Z.N.; investigation, T.Z.N.; writing—original draft, T.Z.N.; writing—review and editing, M.M. and A.B.B.; funding acquisition, A.B.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vaidya, B.; Makrakis, D.; Mouftah, H.T. Multi-domain Public key infrastructure for Vehicle-to-Grid network. In Proceedings of the MILCOM 2015—2015 IEEE Military Communications Conference, Tampa, FL, USA, 26–28 October 2015.
2. Basnet, M.; Ali, M.H. Deep Learning-Based Intrusion Detection System for Electric Vehicle Charging Station. In Proceedings of the 2020 IEEE 2nd International Conference on Smart Power and Internet Energy Systems (SPIES), Bangkok, Thailand, 2–4 June 2020.
3. Tait, K.A.; Khan, J.S.; Alqahtani, F.; Shah, A.A.; Khan, F.A.; Rehman, M.U.; Boulila, W.; Ahmad, J. Intrusion Detection using Machine Learning Techniques: An Experimental Comparison. In Proceedings of the 2021 IEEE International Congress of Advanced Technology and Engineering (ICOTEN), Virtual, 4–5 July 2021.
4. Vinayakumar, R.; Alazab, M.; Soman, K.P.; Poornachandran, P.; Al-Nemrat, A.; Venkatraman, S. Deep Learning Approach for Intelligent Intrusion Detection System. *IEEE Access* **2019**, *7*, 41525–41550. [CrossRef]
5. Available online: https://theicct.org/sites/default/files/publications/charging-up-america-jul2021.pdf (accessed on 22 February 2022).
6. Available online: https://www.nrcan.gc.ca/energy-efficiency/transportation-alternative-fuels/zero-emission-vehicle-infrastructure-program/21876 (accessed on 22 February 2022).
7. Recommendation ITU-T X.1372. Security Guidelines for Vehicle-to-Everything (V2X) Communication; International Telecommunication Union (ITU). Available online: https://www.itu.int/rec/T-REC-X.1372-202003-I (accessed on 22 February 2022).
8. Hoang, D.T.; Wang, P.; Niyato, D.; Hossain, E. Charging and Discharging of Plug-In Electric Vehicles (PEVs) in Vehicle-to-Grid (V2G) Systems: A Cyber Insurance-Based Model. *IEEE Access* **2017**, *5*, 732–754. [CrossRef]
9. Saxena, N.; Grijalva, S.; Chukwuka, V.; Vasilakos, A.V. Network Security and Privacy Challenges in Smart Vehicle-to-Grid. *IEEE Wirel. Commun.* **2017**, *24*, 88–98. [CrossRef]
10. *ISO 15118-1:2013*; Road Vehicles—Vehicle to Grid Communication Interface—Part 1: General Information and Use-Case Definition. International Organization for Standardization: Geneva, Switzerland, 2013.
11. *ISO 15118-2:2014*; Road Vehicles—Vehicle-to-Grid Communication Interface—Part 2: Network and Application Protocol Requirements. International Organization for Standardization: Geneva, Switzerland, 2014.
12. *ISO 15118-3:2015*; Road Vehicles—Vehicle to Grid Communication Interface—Part 3: Physical and Data Link Layer Requirements. International Organization for Standardization: Geneva, Switzerland, 2015.
13. Deepthi, H.L.; Philips, J.; Tabrizi, N. A Survey of Intrusion Detection Techniques. In Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019.
14. Toupas, P.; Chamou, D.; Giannoutakis, K.M.; Drosou, A.; Tzovaras, D. An Intrusion Detection System for Multi-Class Classification based on Deep Neural Networks. In Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019.
15. Bensaber, B.A.; Caroly Gabriela, P.D.; Lahrouni, Y. Design and modeling an Adaptive Neuro-Diffuse System (ANFIS) for the prediction of a security index in VANET. *J. Comput. Sci.* **2020**, *47*, 101234. [CrossRef]
16. Rupareliya, J.; Vithlani, S.; Gohel, C. Securing VANET by preventing attacker node using Watchdog and Bayesian Network Theory. In Proceedings of the 7th International Conference on Communication, Computing and Virtualization, Mumbai, India, 26–27 February 2016.
17. Rampaul, D.; Patial, R.K.; Kumar, D. Detection of DoS Attack in VANETs. *Indian J. Sci. Technol.* **2016**, *9*. [CrossRef]
18. Alshammari, A.; Zohdy, M.A.; Debnath, D.; Corser, G. Classification Approach for Intrusion Detection in Vehicle Systems. *Wirel. Eng. Technol.* **2018**, *9*, 79–94. [CrossRef]
19. Lim, K.; Islam, T.; Kim, H.; Joung, J. A Sybil Attack Detection Scheme based on ADAS Sensors for Vehicular Networks. In Proceedings of the 2020 IEEE 17th Annual Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, USA, 10–13 January 2020.
20. HorváTh, G.; KováCs, E.; Molontay, R.; NováCzkI, S. Copula-based anomaly scoring and localization for large-scale, high-dimensional continuous data. *Acm Trans. Intell. Syst. Technol.* **2020**, *11*, 26. [CrossRef]
21. Lahrouni, Y.; Pereira, C.; Boucif, A.B. Using Mathematical Methods against DeniaI of Service (DoS) Attacks in VANET. In Proceedings of the 15th ACM International Symposium on Mobility Management and Wireless Access, Miami, FL, USA, 21–25 November 2017.
22. Mesfioui, M.; Bouezmarni, T.; Belalia, M. Copula-based link functions in binary regression models. In Proceedings of the 48th Annual Meeting of the Statistical Society of Canada, Virtual, 6–9 June 2021.
23. Brechmann, E.C.; Schepsmeier, U. Modeling dependence with C- and D-Vine Copulas: The R package CDVine. *J. Stat. Softw.* **2013**, *52*, 1–27. [CrossRef]
24. Attanasio, L.; Conti, M.; Donadel, D.; Turrin, F. MiniV2G: An Electric Vehicle Charging Emulator. In Proceedings of the 7th ACM on Cyber-Physical System Security Workshop, Hong Kong, China, 7 June 2021.
25. Available online: https://www.unb.ca/cic/datasets/ids-2018.html (accessed on 22 February 2022).

Chapitre 7

RÉSULTATS

Dans ce chapitre, nous présentons les résultats obtenus suite à la prédiction de la variable binaire "ATT" (attaque abrégée) avec notre modèle de détection d'intrusions basé sur les copules. Les variables explicatives de notre modèles sont les quatre variables significatives déterminées plus haut.

Il est important de préciser que notre modèle a été développé sous le logiciel R et que les informations contenues dans notre base de données V2G générée sont réparties en deux (02) parties. La première partie contient 80% des données pour l'entrainement et l'autre partie est composée de 20% de données tests.

Les résultats obtenus sont les suivants :

— un taux de prédiction de 96,43% ce qui équivaut à un taux d'erreur de 3,57%

— la matrice de confusion de notre modèle se présente comme suit :

		Valeurs réelles		
		0	**1**	**Total**
Valeurs prédites	**0**	38 (TN)	0 (FN)	38
	1	2 (FP)	16 (TP)	18
	Total	40	16	56

TP : True Positive ; TN : True Negative ; FP : False Positive ; FN : False Negative

FIGURE 7.1 – Matrice de confusion relative au réseau V2G

— Recall = $\frac{TP}{TP+FN} = \frac{16}{16+0} = 1$

Ce résultat signifie qu'au niveau de notre échantillon de données tests, notre algorithme a pu prédire la totalité des données relatives aux attaques.

— Precision = $\frac{TP}{TP+FP} = \frac{16}{16+2} = 0{,}88$

La valeur 0,88 signifie que 88% des données qui sont prédites comme des attaques sont en réalité des attaques.

— Le F-measure est considéré comme une combinaison du Recall et de la Precision. Il est alors facultatif dans notre cas, néanmoins nous l'avons calculé :

F-measure = $2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{1 \times 0{,}88}{1+0{,}88} = 0{,}93$

De ces différents résultats obtenus, nous retenons que notre modèle de détection d'intrusions basé sur les copules a donné des résultats très satisfaisants. Ces résultats obtenus sont confirmés lors de l'évaluation de notre modèle avec une autre base de données générée par les auteurs de [21]. Il s'agit d'une base de données d'attaques générée à partir des données collectées dans le réseau véhiculaire VANET.

Les résultats obtenus après l'évaluation de notre modèle avec la base de données VANET sont les suivants :

— un taux de prédiction de 95,83% ce qui équivaut à un taux d'erreur de 4,17%

— la matrice de confusion de notre modèle se présente comme suit :

		Valeurs réelles		
		0	**1**	**Total**
Valeurs prédites	**0**	38 (TN)	0 (FN)	38
	1	2 (FP)	8 (TP)	10
	Total	40	8	48

FIGURE 7.2 – Matrice de confusion relative au réseau VANET

— Recall = $\frac{TP}{TP+FN} = \frac{8}{8+0} = 1$

— Precision = $\frac{TP}{TP+FP} = \frac{8}{8+2} = 0{,}8$

— F-measure = $2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{1 \times 0{,}8}{1+0{,}8} = 0{,}88$

7.1 Conclusion

De l'analyse des différents résultats obtenus, nous retenons que notre modèle basé sur les copules est plus performant dans le réseau V2G que dans le réseau VANET. Ce qui est tout à fait normal puisque les voitures dans les réseaux VANET se déplacent à des vitesses différentes.

Chapitre 8

CONCLUSION GENERALE

Dans le cadre de la protection de l'environnement les véhicules électriques constituent une alternative aux véhicules à essence. Pour faciliter l'utilisation des véhicules électriques il est important de penser à la prolifération des bornes de recharge et à la protection de l'infrastructure Vehicle-to-Grid contre les cybercriminels. Afin de relever le défi en matière de cybersécurité dans le V2G, il est primordial de concevoir des solutions de détection d'intrusions innovantes capables de répondre aux exigences comme la confidentialité, l'intégrité, la disponibilité et la non-répudiation. Ainsi, une forte adhésion de la population aux véhicules électriques sera observée.

Le modèle de prédiction d'attaques proposé pour le réseau Vehicle-to-Grid dans le cadre de notre travail, est basé sur la notion des copules. Ce modèle est inspiré des travaux de Mhamed Mesfioui et autres [30]. Les différentes phases de réalisation de notre modèle se décompose en deux grandes étapes à savoir : la génération de la base de données d'attaques V2G suite aux différentes simulations et la conception de l'algorithme responsable de la prédiction. La capacité du modèle proposé à prédire la variable binaire «ATT» a été évaluée en utilisant la matrice de confusion. Il ressort de cette évaluation que le taux de prédiction de notre modèle est de 96,43%.

Notre sujet de recherche, nous a permis de mettre en évidence les attaques de types déni de service et homme du milieu grâce aux différentes simulations effectuées. Aussi, nous avons pu prédire une attaque qu'elle soit de type DoS ou MitM à l'aide de la méthode de prédiction basée sur les copules. Dans la continuité de la présente étude, il serait intéressant de travailler sur notre modèle pour le rendre capable de déjouer les

attaques par déni de service distribuées. Il est aussi envisageable de localiser la source d'une attaque dans le réseau afin de réduire le temps de réponse suite à un incident.

Bibliographie

[1] ZEGHMAR Radia, ABDERRAHMANE Nour El Houda :*Modélisation De La Dépendance Par Les Copules* , 2016.

[2] Nabil KADI : *Estimation non-paramétrique de la distribution et densité de copules*, faculté des sciences Université de Sherbrooke, Canada, 2014.

[3] Bezat A. et Nikeghbali A. :*La théorie des extrêmes et la gestion des risques de marché*, Mai 2000.

[4] Nelson, R. B. : *An Introduction to Copulas.* Springer Series in Statistics, Springer Science+Business Media, Inc., New York, 2nd édition, 2006.

[5] Lounas Fadhila : *Modélisation de la dépendance par les copules et applications*, Université de Mouloud Mammeri, Tizi-Ouzou, Algérie, 2011.

[6] Salvadori, G., Michele, C. ,Kottegoda , N. T. and Rosso, R. : *Extremes In Nature : An Approach Using Copulas,* ,Springer, The Netherlands, 2007.

[7] Gumbel, E. J. : *Bivariate logistic distributions* , 56 : 335–349, 1961.

[8] GILLIS DELMAS TCHOUANGUE DINKOU : *Le modèle de régression quantile binaire à base d'une copule*, Université du Québec à Montréal, Canada, 2020.

[9] Malevergne, Y. et Sornette, D. : *Extreme Financial Risks* , From Dependence to Risk Management, Springer Berlin Heidelberg, New York, 2006.

[10] Hult, H. and Lindskog, F. : *Mathematical Modeling and Statistical Methods for Risk Management* , France, 2002.

[11] Nesrine Idiou, Fatah Benatia, Mounir Mesbah : *Copules et modèles avec variable de fragilité pour des données de survie multivariée*, 2021.

[12] Franke, J., Hardle, W. K. et Hafner, C. M. : *Statistics of Financial Markets An Introduction,* Springer-Verlag Berlin Heidelberg, Second Edition, 2008.

[13] SAIDOU DIOP. : *Une Infrastructure à Clés Publiques (PKI) pour sécuriser les messages dans un réseau V2G* , Mars 2018.

[14] `https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions`

[15] `https://theicct.org/sites/default/files/publications/charging-up-america-jul2021.pdf`

[16] Recommendation ITU-T X.1372, *Security guidelines for vehicle-to-everything (V2X) communication*, 2020.

[17] Dinh Thai Hoang, Ping Wang, Dusit Niyato et Ekram Hossain, *Charging and Discharging of Plug-In Electric Vehicles (PEVs) in Vehicle-to-Grid (V2G) Systems : A Cyber Insurance-Based Model*, IEEE Access, 2017.

[18] Neetesh Saxena, Santiago Grijalva, Victor Chukwuka et Athanasios V. Vasilakos :*Network Security and Privacy Challenges in Smart Vehicle-to-Grid*, IEEE Wireless Communications, IEEE, 2017.

[19] Deepthi Hassan L. , James Philips , Nasseh Tabrizi, *A Survey of Intrusion Detection Techniques* , 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), IEEE, 2019.

[20] Petros Toupas , Dimitra Chamou , Konstantinos M. Giannoutakis , Anastasios Drosou, Dimitrios Tzovaras : *An Intrusion Detection System for Multi-Class Classification based on Deep Neural Networks*, 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), IEEE, 2019.

[21] Caroly Gabriela P. D., Boucif Amar Bensaber , Youssef Lahrouni, *Design and modeling an Adaptive Neuro-Diffuse System (ANFIS) for the prediction of a security index in VANET*, 2019 IEEE Symposium on Computers and Communications (ISCC), IEEE, 2019.

[22] Jay Rupareliya, Sunil Vithlani, Chirag Gohel, *Securing VANET by preventing attacker node using Watchdog and Bayesian Network Theory*, 7th In-

ternational Conference on Communication, Computing and Virtualization, Elsevier B.V, 2016.

[23] Deepak Rampaul, Rajeev Kumar Patial, Dilip Kumar :*Detection of DoS Attack in VANETs*, Indian Journal of Science and Technology, 2016.

[24] Abdulaziz Alshammari, Mohamed A. Zohdy, Debatosh Debnath, George Corser :*Classification Approach for Intrusion Detection in Vehicle Systems* , Wireless Engineering and Technology, Scientific Research Publishing, 2018.

[25] Kiho Lim, Tariqul Islam, Hyunbum Kim, Jingon Joung :*A Sybil Attack Detection Scheme based on ADAS Sensors for Vehicular Networks*, 2020 IEEE 17th Annual Consumer Communications and Networking Conference (CCNC), IEEE, 2020.

[26] GÁBOR HORVÁTH, EDITH KOVÁCS, ROLAND MOLONTAY, SZABOLCS NOVÁCZKI :*Copula-based anomaly scoring and localization for large-scale, high-dimensional continuous data* , ACM Transactions on Intelligent Systems and Technology (TIST) - Survey Paper and Regular Papers, 2020.

[27] Youssef Lahrouni, Caroly Pereira, Amar Bensaber Boucif :*Using Mathematical Methods against DeniaI of Service (DoS) Attacks in VANET*, 15th ACM International Symposium on Mobility Management and Wireless Access, 2017.

[28] Luca Attanasio, Mauro Conti, Denis Donadel, Federico Turrin :*MiniV2G : An Electric Vehicle Charging Emulator*, 2021.

[29] `https://www.unb.ca/cic/datasets/ids-2018.html`.

[30] Mhamed Mesfioui, Taoufik Bouezmarnib et Mohamed Belaliac :*Copula-based link functions in binary regression models* , 48th Annual Meeting of the Statistical Society of Canada, 2021.

[31] Eike Christian Brechmann, Ulf Schepsmeier :*Modeling dependence with C- and D-Vine Copulas : The R package CDVine*, Journal of Statistical Software, 2013.

[32] ISO 15118-1 :2013 ; International Organization for Standardization, 2013, Road vehicles — Vehicle to grid communication interface — Part 1 : General information and use-case definition.

[33] ISO 15118-2 :2014 ; International Organization for Standardization, 2014, Road vehicles — Vehicle-to-Grid Communication Interface — Part 2 : Network and application protocol requirements.

[34] ISO 15118-3 :2015 ; International Organization for Standardization, 2015, Road vehicles — Vehicle to grid communication interface — Part 3 : Physical and data link layer requirements.

[35] Nonvignon, T.Z. ; Boucif, A.B. ; Mhamed, M. A Copula-Based Attack Prediction Model for Vehicle-to-Grid Networks ; Appl. Sci. 2022, 12, 3830.