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.

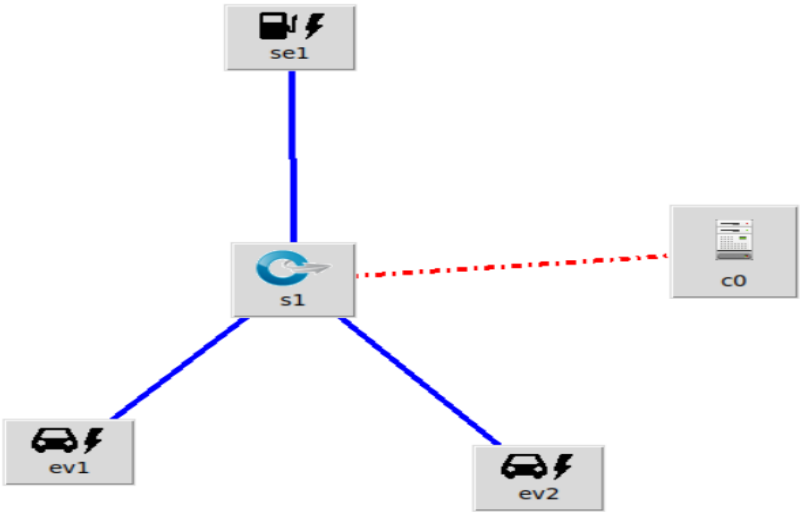

Figure 1. Attack-free simulation architecture.

(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).

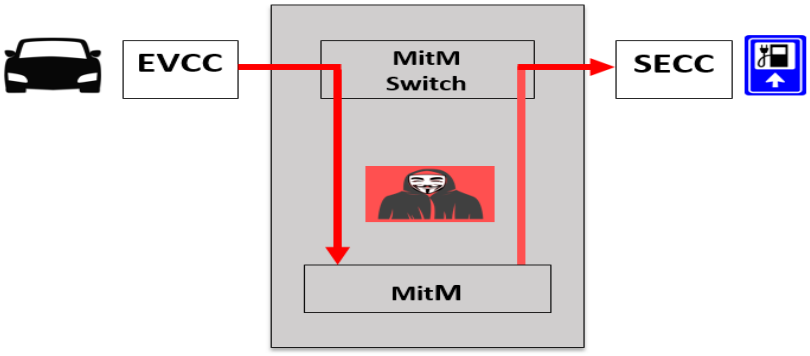

Figure 2. Architecture relating to the simulation of the MitM-type attack.

(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).