SURVEY ON MALWARE DETECTION APPROACHES IN VANET

BY

MUHAMMAD ANWAR SHAHID

DRAFT ONLY

ABSTRACT

VANETs represent a network infrastructure that allows for wireless communication among vehicles on the road. Consequently, enhancing comfort, safety and convenience. Due to its unique characteristics, ad-hoc networks and VANETs are different. However, VANETs are prone to misbehaviors resulting from inadequate centralized administration and related deficiencies; and this affects VANETs security in different ways. An attacker node can create a fake message or force another node to generate a different node. The paper shows a ML (machine learning) method to show misbehaviors portrayed by VANETs. Also, the paper identifies major security issues encountered in VANETs while exploring various Machine learning approaches to develop solutions for major security and privacy issues. Ultimately, the paper presents key areas for future research in the field of VANETs with regards to major machine learning approaches.

Table of Contents

Chapter1: Introduction to VANETs

In Vehicular Ad Hoc Network (VANET), small packets are occasionally swapped between nodes if in the area to tighten up safety when driving. Most applications by VANET aim to improve safety for drivers. VANET security represents an essential issue because applications with harmful nodes may broadcast and produce wrong or incorrect text leading to unwanted effects in the network. During periodic vehicle movement safety packets transmit between nodes in a specific area within VANETs. The most significant function of VANET applications involves the enhancement of safety for drivers. Nevertheless, undesirable consequences may occur due to malicious nodes, which may lead to inaccurate propagation of messages. Ultimately, VANET security represents an essential factor in numerous application.

Various experiments were done to derive structures from packet transmission and physical features related to original and malicious nodes. In this regard, a study by Kamel et.al. (2020) attempts to identify misbehaving nodes by utilizing data and suggests algorithms that identify fabricated alert messages by monitoring node behavior. Legitimate or malicious nodes are classified by the use of machine learning approach classifier. The accuracy is largely affected by factors below;

(a) Naive Bayes, Artificial decision tree and Neural network are some of the inducer logarithm used in generation of a classifier.

(b)Features like representation make each classifier work in a different way when produced using various structures and inducers.

Moreover, it is wise put in to use a number of classifiers in order to put up a node as genuine or harmful then put together the consequences of all the classifiers by means of a good combination aproach into one final outcome. This type of classification becomes more accurate in contrast to correctness achieved through individual classifiers this helps find out more about its weaknesses and strengths of each classifier. This implies that using many diverse classifiers to organize a node as either malicious or legitimate allows for the integration of the outcomes into a final result via an intelligent combination method.

In the best of our knowledge, this method has not been used earlier to differentiate misbehaviors in VANETs. The most substantial contribution in this paper entails the implementation of misbehavior detection and design on the basis of ensemble learning. Additionally, `ensemble based learning results' are done in comparison with binary classes.

VANET Network Architecture

Communication between roadside and vehicles elements are made possible by VANETs which represents a distinct type of `mobile hoc network (Gyawali, & Qian, 2019, pp. 1-6). Traffic management and congestion monitoring are some of the emerging technologies that are facilitated by VANETs. i.e. when an accident occurs on a particular road, each vehicle is notified through an alert to divert routes in a bid to avoid traffic congestion arising from the accident. Moreover, other applications of VANETs include payment services, infotainment, and premium calculations.

However, improvements are in progress to benefit residents in smart cities. For instance, the complexity involved in managing transport and logistics, makes it better and smarter to fulfill the daily expectations of many citizens. This implies that intelligent vehicles are expected to change the future transport system. Therefore, manufacturers are trying their best to upgrade vehicle independence levels in present production lines to automate various driver functions. Also, intelligent transport systems attempt to incorporate functions to collect and exchange bulk data while on high speeds to enhance timely decision making (Haydari, and Yilmaz, 2018 pp. 157-163). The integration of the new generational vehicles with; Event Data Recorder, On-board Units, Global positioning system, and other essential sensors improves the chances of success. The aforementioned devices allow for the detection of traffic conditions, while subsequently sharing the acquired data through V21 (Vehicle to Structure) or V2V (Vehicle to Vehicle) within the network. In this regard, Dedicated Short Range Communication (DSRC) technology is used in the ITS (Intelligent Transportation Systems) area to facilitate for dependable and safe links of communication in both V2I and V2V approaches

Vehicular ad-hoc Network represents a distinct type of MANET which allows for communication between different vehicles and installed devices along streets and highways (Biswas, Mišić, and Mišić, 2012 pp. 1079-1084). Also, nodes help in enabling communication. There exists two kinds of nodes including;

(a) The infrastructure and mechanisms relative to MANET networks erected along the highways and streets, commonly known as RSU.

(b) The devices installed in vehicles commonly referred to as OBUs.

Types of Communication in A VANET Network

V2V: Valuable information on the road such as ambulance passage or an accident is shared between vehicles.

V2I: Communication between network infrastructure and vehicles, the RSU helps to connect to the external network (the internet).

I2V: Different situations in traffic and along street are shared by sending messages to vehicles in motion.

Characteristics of VANET Network

i. High mobility of nodes: This is a crucial factor in VANETs because nodes continuously move at different speeds and in diverse directions.

ii. Active topology: VANET is dynamic.

iii. Frequent network interruptions: Network disconnection is mostly caused by; traffic jams, node flexibility, climate change and dynamic topology.

iv. Time constraints: Time denotes a crucial factor in message delivery because most applications need to receive messages at a specific point in time.

v. Unlimited battery power and storage: This network has sufficient storage.

vi. Unbounded network size: There exists no geographical limitations in VANET networks thus it can be installed across a single city, a region or the whole country.

VANETs Security Framework

For VANET's security and its applications, more so applications involved in safety, it is vital to verify transmitted communications and qualities of their despatchers. Failure to do so would lead to the distribution of bogus messages among random unauthorized vehicle while also engaging in other unnoticeable mischievous activities. These activities would result in inordinate damages towards urban transportation systems and even subject the lives of drivers and pedestrians in danger. Existing safety necessities that must be considered in developing safe planning for VANET include;

- Its Authentication: Initially, the first and most important step for vehicular network security is indeed authentication. It is vital for confirming an entitlement of validity. Specifically, in VANETs authentication refers to validating the vehicle's identity and differentiating genuine vehicles from illegitimate vehicles (Raw, Kumar, and Singh, 2013 p.95).

- Accountability: This is where the node involved in message transfer is bound to its activities. Law enforcement agencies should identify various malicious drivers and hold them accountable for their activities.

- Data Consistency: This commonly engages usability, authenticity, accuracy, and data integrity in the vehicular networks. Also, it makes sure that every driver in the system uses reliable data to navigate through traffic.

- Integrity: A wireless channel is venerable to attacks, such as, alteration of data. Integrity assures that the messages do not succumb due to the attacks and that all sent messages do not undergo alteration.

- Non-repudiation: This entails thwarting the denial that a message has been received from a specific entity. Therefore, non-repudiation remains a vital condition for the unfailing use of VCS.

- Restricted Credential Usage: To attain both accountability and authentication, cryptography is used in form of a token referred to as a credential. The limitation of matching the use of credentials in authentication at certain times entails a critical requirement for security.

- Credential Revocation: Because VCS attaches the element credential of trust to a node, there exist the necessity to establish a method to nullify a credential. In the case of ill-functioning or defective nodes, removal of the nodes from the network is required and it is initiated via the annulling of their credentials.

To achieve the aforementioned security requirements, the following frameworks are engaged including;

Public Key Infrastructure (PKI) based schemes. This provides a multitude of public/private pairs of keys and certificates for vehicles. A digital signature and a certificate ought to be attached before sending a message, this could lead to a significant upsurge in the communication overhead. To realize privacy of identity and provisional anonymity, unsigned public keys ought to be imposed for automobiles and PKI. The certificate management as well as annulment would be a huge encumbrance to PKI.

Symmetric cryptosystem based schemes. (MAC) Message authentication code could be engaged in message authentication, while effecting the verification of messages to finalize in an instant. Conversely, the course of verifying messages could require the assistance of RSUs, besides vehicles cannot exclusively authenticate the message received. TESLA remains an effective protocol for broadcast authentication centered on time synchronization and MAC loss amid nodes in the network. In the context of TESLA, vehicle direction prediction can be possible to promptly verify beacon messages sent by automobiles. Regrettably, the protocol allows for easy traceability of vehicle trajectory by attackers.

Schemes based on group signature: This obviously provides discretion to members in the group since any associate is allowed to sign the message rather than the whole group. The management in charge of the group possess the group master key and is capable of learning the actual group members' identities, which fulfills the conditional privacy preservation requirement. Though, the group signature verification typically results in more time costs compared to the customary signature. Similarly, it remains problematic when correctly annulling group members that are compromised.

Schemes based on IBC. For the (IBS) identity-based signature scheme, the public key can be used as the vehicle, and the matching private key is produced by the (PKG) private key generator that uses the master key. Relating to PKI, certificates management is avoided. For conditional privacy to be realized, vehicles link to entities by means of pseudo-identities which can be retrieved by authorities. Regrettably, IBS schemes time efficiency is moderately slower compared to customary signature schemes due to bilinear pairing operations.

Scheme based on Pseudonymous Authentication. Every vehicle stocks numerous pseudonymous certificates initially, and afterward randomly selects one of the certificates to act as the identity at a certain time. Likewise, since a Trusted Authority (TA) has adequate storage and cannot

be conceded, it is feasible and safe for a TA to stock the pseudonymous certificates. Once a vehicle originally lists, the TA directs adequate pseudonymous certificates to it, and a unique permanent identity. For privacy concerns, vehicles do not engage the permanent identity to sign messages; they instead indiscriminately select one of the pseudonymous certificates that the TA has provided for the digital signature. By this method, the short-term identity for every vehicle varies over time, and a mischievous attacker is unable to trace a precise vehicle. This is because after modifying the certificate, the attacker shall be unable to relate the old certificate with the new certificate, which translates that the attacker has lost the objective. Nonetheless, this technique still experiences various complications, such as great revocation costs. For example, when there is a revocation of a vehicle, the number of pseudonymous certificates thought to be supplementary to the Certificate Revocation List (CRL) might be too vast, whereas the CRL mass rises quickly when the network mass rises.

Chapter2: Security Threats in VANETs

A lot of security issues have developed with the utilization of the wireless ad hoc networks. Some of the security issues that have emerged and need much attention include: lack of central point, lack of a clear line of protection, wireless cooperativeness and mobility. With VANETs, the mentioned issues are made more complex. In addition to those issues, VANETs have more challenges which are discussed in details below

Privacy: It is rather hard to offer both privacy and security at the same time. Taking an example of the necessity of the authorities to have vehicle drivers' data for provision of security while the vehicle owner prefers to be discrete with his or her personal information like their identity and the history of location or transactions. This situation explains best how it is hard to offer both security and privacy at the same time. In a bid to prevent all vehicle tracking like in the incidence of 'big brother', a system should prioritize user's privacy while permitting for establishment of the user's real identity once requested by legit authorities such as the police, court or the manufacturers. (Rivas et.al, 2011, pp.1942-1955).

Scalability: The number of vehicles on the global roads is approximated at over one billion and is still rising. Among the over 1 billion, those that have a connection to VANETs are over 250 million as recorded in 2020. Currently, no global authority can that offer security of such networks. This is because it's very difficult to lay out standardized regulations for VANETs; a problem brought about by the previously mentioned security-privacy trade-off differs greatly from one state republic to another (Jeevitha, and Bhuvaneswari, 2019, pp.1-6). There should be coordination of local authorities globally so that they may be provision of standardized security for the VANETs networks.

Mobility: VANET topology transforms very quickly since it is a one-time link among vehicles. Although the nodes are seen shifting at a reliable speed of 20 meters per second in MANET simulations, the motor vehicles actually go at a much faster speed than the maximum velocity.

As a result, connections between vehicles constantly break; a challenge commonly faced. This is challenge is majorly faced when vehicles are going in opposite directions. The vehicles only stay in communication for a short while and then the network is disconnected. This quite a challenge to the reputation-based systems despite the fact that the topology changes in an expected way than simulations from MANET as indicated in the literature review.

Hard-delay constraints: Majority of the applications in VANETs needs a big-time response. If not so, the results could be lethal like accidents could occur or even delayed rescues. The necessity of real-time responses is what also leaves the systems defenseless to Denial of Service (Dos) occurrences. Researchers suggest that the safety related systems should majorly focus on curbing the attacks instead of putting focus on the process of detecting and recovering because of the demand brought about by the real-time issue (Pathan, 2016).

Cooperativeness: Many of the algorithms and protocol of VANETs observe that information is transferred by vehicles when they are in communication. This aspect makes the VANET susceptible to attacks occurrences like the fake information attacks. Many of the security advancements also rely on the cooperativeness of the motor vehicles since local data may not be enough for the detection and prevention of attacks issues.

Wireless ad hoc networks are made up of many devices that distinguishable storage and computational capacity; ranging from hand-held gadgets to very strong laptops. These devices majorly run on power from batteries. For that reason, MANETs face a challenge of limited resources. This is a challenge that must be looked into while in search of solutions for the security issue. VANETs nodes can be hard target for energy exhaustion attacks such as those that attack MANETs; Sleep Derivation Torture since they have sufficient power and are stable roadside units (RSU) (Biswas, Mišić, and Mišić, 2012 pp. 1079-1084).

Absence of clear line of defense is a challenge that is faced by MANETs but has less impact on the VANETs. Regarding the aspect of VANET is lacking central points on which security mechanisms should be positioned on, wired networks can play an important role of transferring resource intensive communications like those of collecting alarms that have been raised by vehicles and making a clear decision. These applications and motor vehicles unlike MANETs, can be more secured and protected. RSU-based security resolutions such as RSU based intrusion detection and RSU-aided certificate revocation are already being projected in the literature as well as large scale positioning of RSUs is a costly method.

Types of Attacks

In consideration of conception of VANETs, new security challenges have risen with time. Although this was anticipated in the networks, there is the need for a deep understanding of the attacks. In this section, prevailing attacks on VANETs are categorized based on their methods and goals. The specific attacks to most routing protocols may not be stated in the literature review since are not going

to be put in consideration in the future study. Although some of the occurrences are sourced from prevailing attacks on MANETs, majority of all the explicit to VANETs.

Sybil attack

This occurrence may be categorized since one VANETs dreadful attack. In such an attack situation, a node (vehicle) can act offering more than a single identity. In other terms, the rest of nodes encrypted in the network are have no ability to differentiate if the data sourced from one motor vehicle or from more than one motor vehicle. The key purpose done by the attacker is shaping the networks rendering his/her objectives. For instance, an attacker has the potential to manipulate other motor vehicle‟ behaviors for example having them take another road rather than their planned route. Sybil attack is also amongst the hardest attack to identify apart from it being most harmful of kind of attack (Li, and Zhang, 2019 pp. 763-767). The attackers perceive it as the most dreadful attack, since it uses the geographic location to find the find the location of the automobile.

Additionally, this may indicate occurrence happening in locations that isn't legit location. The common type of Sybil attack is known as *Node Personation Attack*. Since every vehicle in the network has its distinct identity, the unique feature is used by the cars to communicate among each other. (Li, and Zhang, 2019 pp. 763-767). In cases where vehicles might change their identity and the system does not recognize the change, an attack is launched on the vehicle since the system usually don't recognize the system. In cases where a vehicle ais caught in an accident it might change its identity and seem like a vehicle that is moving hence make recovery hard. This makes all other vehicles in the system to find the vehicle in a way that the vehicle is moving while actually is it on the stand still in an accident. Then, the malicious vehicle might convey inappropriate info concerning the road situation to the area around RSUs.

Denial of service breakdown

In this attack, the core emphasis is the accessibility of the valid actions of a system. The attackers normally request for actions that are far way too many for the system to cope with. In VANETs, an assailant tries to put down the network by introducing RSUs and stopping communication between motor vehicles. As a result of DoS assaults, the motor vehicles are unable to receive and send information on the network and thus leads to fatal incidents (Alheeti, Gruebler, and McDonald-Maier, 2015 pp. 86-91). Nodes may establish an assault from several areas in the case of Distributed Denial-of-Service assault. This property makes any detection very hard. Nodes that establish a DDoS assault could major in not only damaging of the network's vehicles but also the system of RSUs.

There are many kinds of DoS assaults on VANETs. The common type of attacks discussed in this literature include:

Jellyfish attacks.

According to Alheeti, Gruebler, and McDonald-Maier, (2015 pp. 86-91) Jellyfish assault is generally a MANET's protocol-complaint DoS assault. Unlike many of the other routing assaults, this follows each and every routing procedures. The assailant could basically delay, disorient or systematically drop packages which they were expected to forward. Ultimately, it cultivates amenability of end to end control conventions so as to radically bring down the performance of the network. VANETs could easily inherit the Jellyfish assaults.

Intelligent cheater attack has no much difference from the Jellyfish attack considering that it also follows the specifications of routing conventions (Pathan, 2016). The assailant discontinuously misbehaves as the operations appear to be normal many of the times. In that manner, the attacks can easily bypass trust mechanisms. It becomes rather hard to detect these type of assaults because of their sneaky nature and for that matter there is the need for end to end control mechanism in addition to a lasting monitoring. However, the mobile nature of the VANETs makes the idea of long term monitoring a very hard practice.

Flooding assaults on the other hand, create traffic in the networks so as to deplete the resources like power, bandwidth and the CPU. These attacks can be categorized into two: routing control packets flooding and data flooding. Despite their different grouping, the consequences of the two types of flooding attacks are similar. In both, the genuine users of this network are limited to the resources. The difference come in that a data flooding assailant generates useless information packets to all nodes by means of the neighbors while the other, the attacker sends route request control packets to network nodes that do not even exist. In the data flooding assault, there is the need to establish routes with all possible nodes prior to the attack of the network.

Jamming attack is another DoS assault. It entails the occupation of the network channel by generating radio frequencies that comprise of illegal traffic. The attack can be carried out by an individual who is not even a member of the specific network. Putting into consideration that any individual with even little knowledge on VANETs could carry out DoS assaults and cause vehicles to be left out in real life traffic, the probability of these attacks being performed is very high. It is necessary that these attacks be detected as early as possible a response action be activated in a timely manner because is notably hard for the network to respond when under a successful attack. Apart from

the mentioned actions, there is need for application of mitigation approaches as those presented by Biswas, Mišić, and Mišić, (2012, pp. 1079-1084).

Black-hole Attack

Unlike the DDoS and the DoS whose objective is to close down the network, a blackhole attack shapes the network. It a threat of high magnitude to the MANETs. The assailant is manipulative and redirects nodes to packet dispatching through itself. An assailant in VANETs can utilize routing conventions like convincing that it knows the suitable way for the motor vehicle's destination or it is the excellent path to dispatch the packages. By generating wrong routing data, it convinces vehicles to dispatch their packets through itself. After this misrouting, the vehicles which have fallen victim dispatch their packages to the assailant, it intentionally rejects the packets and finally there are losses of packets in the entire network. The figure below is an illustration of the same.

Alheeti, Gruebler, and McDonald-Maier, (2015 pp. 86-91)presented an analysis done in, it is proven that black hole assaults in VANETs impact the networks in manners such as the end to end hindrance for network loading leading to the AODV being susceptible to the assaults than the OLSR. Apart from the simple act of losing the packets, the assailants can also transfer the packets between each other in a manner that they are making a network of their own. An instance is when route request reaches a malicious vehicle, it can transfer the request to yet another motor vehicle of malice. Through this, network's crucial information is not promoted by the assailant and may not be received by other vehicles since a distractive motor vehicle creates a communication only between themselves and neglects the rest of the network members. This implies that messages of safety cannot be communicated to vehicles that are not malicious.

Wormhole Attacker

This attack is majorly carried out by two or more combined nodes which indulge themselves in several routes by showing off that they are aware of the shortest way to any location. The assailant's aim is to reshape the network's logical topology so as to manipulate and bring together big magnitudes of traffic. So as to carry out the assault in VANETs, on receiving a package which should otherwise be forwarded, an assailant motor vehicle confines it and dispatches it the next malicious vehicle. This is followed by the latter distributing the open packet (Singh et.al. 2019 pp. 651-661). Just like the in the black hole assault, the two compromised vehicles convinces the routing convention that the connection between them is the best one to a location over the existing real routes in the network. It is worth noting that the assault can be carried out with the least compromising of a node. This can be done by the

assailant recording the traffic at an instance and then channel them to another through an out-of-band tunnel so that it may replay or reuse it in another instance. Ultimately, crucial information dispatched via the tunnel is not broadcasted leading to a break in communication. In other incidents, the vehicles can make a private network of their own.

Attacks on integrity and data trust

Vehicles in a VANET connection, utilize data which is forwarded or dispatched by the other vehicles. Nevertheless, it is not all the times that the information received is true. A vehicle can just randomly create wrong information and generate it to its network (Ftaimi, and Mazri, 2020 pp.1-8). With selfish intentions, an attacker manipulates the vehicles with false information. An example being the generation of information about a fake accident on a particular road by a malicious vehicle which leads to the rest of the vehicles to take other ways. This form of attack is mainly effective when there is lack of another vehicle to give verification of the information. The impact of this type of attack is more serious is at all it a motorway assailant; one that goes around very fast as it generates false information to groups of vehicles that it meets (Ftaimi, and Mazri, 2020 pp.1-8). Detection of such acts is very hard because each group of vehicles have different networks. There is need for the attacker to have detailed knowledge of the networks so as to prevent use of statistics for their detection. This aspect of the attacks is what minimizes the probability of the attacks.

Illusion Attack

This type of attack is special to the VANETs. The assailant majorly utilizes the intuition of human psychology. To carry it out, the assailant impacts drivers' behaviors by generating false information in accordance to the situations (Almalki, and Song, 2020, p.22). Prior to the attack, the assailant requires to make a conditioned traffic situation. This is because the other drivers are likely to believe the other drivers in the situation once the false information is dispatched. For example, in a situation that the vehicles in the front of the traffic are in a slow motion, by default, the drivers behind them believe that there is an accident as the disseminated information states. They would then turn to use alternative routes on believing the wrong information. For the attack to be perfect, there is the need for the assailant to give out a corresponding wrong information by giving wrong message to its sensors so that the report generated is a valid one despite it being false (Almalki, and Song, 2020, p.22). With message remaining valid, the false information is successfully transmitted to the entire network. Due to that property, this kind of attack is hard to detect.

GPS Spoofing

This type of assault is also referred to as tunnel attack. By utilization of GPS simulators, an assailant can put false position information on other vehicles. The affected vehicle can await for a GPS signal long after it has left a jammed up location or a physical tunnel (So, Sharma, and Petit, 2018 pp. 564-571). The implanted GPS simulator can give out signals that are way much stronger than the rightful GPS signals. Through that, even when a vehicle is in reception of the right satellite signals, it preferably takes in the false position information dispatched by the assailant.

Replay Attack

A replay attack represents a data falsification type of attack, where information is stored for future reuse purposes where it is used to deceive the components of a cITS (Cooperative Intelligent Transport System) (Almalki, and Song, 2020, p.22) including, road side units, vehicles and backend systems. Its main objective being to utilize the situations at the very opportune time of disseminating the creative message. On collecting the information that roams around, the assailant stores the information and later regenerates it in the network despite it being invalid by that time. Surprisingly, the assault can be carried out by the original sender of the message. For instance, an assailant can save a dispatched information on an accident on traffic and later resend it just as illustrated in the figures below. This is easy to do not until the message expires. In that regard, probability of such occurrences is low because of the use of information timestamps to verify messages.

Chapter 3: Introduction to machine learning

Machine learning is artificial intelligence in some way. Understanding the data structure is the objective of machine learning and fitting the data into well-understood replicas for utilization by people. Arthur Samuel created the term machine learning in 1959; he is who one of the American pioneer in the sector of artificial intelligence and computer gaming and did state that “computers are given the ability to learn without having to program specifically. Machine learning is different from traditional computational methods, although it is a computer science field. Traditional computing entails setting algorithms of specifically programed instructions that are used for calculations or problem solving by the computers.

Algorithms for Machine learning allow training of computers on data inputs and output of values that fall in a given range using statistical Analysis (Ftaimi, and Mazri, 2020 pp.1-8). This makes machine learning to facilitate the computers to build models from sample data for automation of processes for decision making basing on data inputs Also, these algorithms come in a variety of styles, with several being published daily and grouped by either the style of learning, i.e. semi-supervised erudition, unsupervised learning, supervised learning or function or form similarity i.e. deep learning, classification, decision tree, regression, and clustering (Taherkhani, and Pierre, 2016, pp.3275-3285). All combinations of learning algorithms in machines comprise of a set of classifiers representation or computer language it understands regardless of the style of learning or function. Consequently, the objective is represented by evaluation. In contrast, optimization involves the search method, the classifier that is the highest scoring often, for example, both are custom optimization method and off the shelf.

Chapter 4: Types of machine learning

Machine learning entails algorithm or computer program teaching to improve on tasks that it is assigned continually. On the side of the research side, machine learning can be defined by considering the mathematical and theoretical models and the working of this process (Grover et.al. 2011 pp. 644-653). There are three main categories of machine learning: ,reinforcement learning, and controlled learning.

Supervised learning

Supervised learning offers a robust tool for processing and classifying data using machine language (Ye et.al. 2018 pp.94-101). Labeled data is used in supervised learning that is a set of data that has been categorized for a learning algorithm to infer. Prediction of the grouping of other unlabeled data by using learning algorithms of the machine is based on the data set. There are two vital techniques in supervised learning which are; classification techniques and linear regression. Linear regression is a type of supervised learning technique that is used for predicting, forecasting, and determining relationships between quantitative data. It is one of the initial methods of learning, which is currently being applied widely. Furthermore, this section will discuss those classification techniques that focus on the prediction of a qualitative responses through data analysis patterns that are recognizable.

The goal of supervised learning methods is developing a group of decision rules to be used for the determination of the known outcome. These can also be named rule induction models, and they comprise the regression models and classification. Algorithms of supervised learning can be used to make rule sets or decision rules, that works by data being subdivided into groups repeatedly basing on the predictor variables identified, that are related to the group membership selected (Ye et.al. 2018 pp.94-101). Besides, a sequence of decision rules created by these techniques can be used for the separation of data into groups that are specific and predetermined. Some model algorithms are developed for particular categories of data, whereas others can only accommodate continuous data. Rule induction models using data that is continuous, although they still parse them to categorizing them via identifying breaks or creating cut points in the given range.

Unsupervised learning

Unsupervised learning algorithms are used for grouping cases depending on their similarity of attributes, or trends that occur naturally, relationships or patterns in the data. The models are also known as self-organizing maps. These Self-organizing maps and clustering techniques are examples

of unsupervised models. Different strategies are used to divide data into groups for different algorithms. Some methods are direct, whereby they divide the cases into groups very fast based on their similarity in attributes or any other common factor. The two-step clustering technique is different in that for its optional number of clusters, a determination is in the first pass through the data, depending on given statistical criteria. The second pass through the data makes the group assignment; therefore, the name “two-step. “Neural networks are more complex than some of the other algorithms of unsupervised learning; consequently, it can give opaque results and interpretation (Ghaleb, Zainal, Rassam, & Mohammed, 2017, pp. 13-18).

The goal of unsupervised learning is to extract regularities from the data sets for their description simplification to reduce the most distinctive elements (Ye et.al. 2018 pp.94-101). Datasets are decomposed into subgroups by clustering, which then gives the initial data summary. Information about present the main trends in the data set is contained in an accurate and simplified description. Still, also, there are typical behaviors and exceptionality coefficients for measuring the extent of representation of subgroups in the entire dataset. An example for illustration is a device that is having three kinds that are explained linguistically as high”, “low” and “abnormally low”: the natural description mentioned the two main cases except for one which is very necessary: it is vital to be aware that the device may be in a state that should not be ignored. Furthermore, “abnormally” adverb emphasizes the point about the three kinds not having similar representation as well as distinguishing the device from the process with three normal modes given as “high,” “low,” and “Very low” for example.

Reinforcement learning

Reinforcement learning involves the training of models of machine learning for making a sequence of decisions. The goal is achieved by the agent in a potentially complex and uncertain environment. Reinforcement learning entails artificial intelligence facing a game-like situation. The computer uses trial and error to generate a solution to a particular problem. To get the programmer do what he wants with the machine, artificial intelligence is either rewarded or penalized for those actions it executes. Its objective is the maximization of the total reward. Reinforcement learning is different from supervised learning in that the training data in supervised learning has the answer key such that the training of the model is done with the correct answer while there is no answer in reinforcement learning (Ye et.al. 2018 pp.94-101). Still, reinforcement agent does the decision on what to do with a specific task. It learns from experience in the case where the training dataset is absent.

The designer does not give the model any suggestions or hint for solving the game though he/she is the one who sets the reward policy, which is the game rules to follow. The model has the liberty to determine the methods of performing the task for maximization of the reward, beginning with total random trials and ending with sophisticated tactics and exceptional human skills. Reinforcement learning is the best way to suggest a machine's creativity by using the power of search and numerous trials. Artificial intelligence can combine experience from thousands of parallel gameplays provided that the reinforcement algorithm is run on adequately powerful computer infrastructure.

There exist two reinforcement types i.e., negative and positive. Positive reinforcement happens due to given behavior either by increasing the frequency and strength of behavior. Negative reinforcement is termed as strengthening behavior due to stoppage of negative condition.

K-means clustering

K-means collecting is an example of unsupervised learning that is applied when there is unlabeled data (i.e., data lacking distinctive groups or categories). The purpose of this algorithm is to discovery groups that occupies a certain data and representation of the identified groups using variable K. Every data point is allocated one of K groups by the algorithm based on the provided features. The similarity feature is the parameter used to cluster data points. Changing from the traditional methods, Taherkhani, and Pierre, (2016 pp.3275-3285) developed an efficient machine learning by basing on control of data congestion tactic that utilizes clustering of k-means for intersections of congestion prone. The suggested tactic depends on local roadside units (RSUs) installation at every connection for recognition of congestion, processing data, and control of congestion for the provision of central congestion supervision for all passing vehicles through or stopping at the connection. After congestion detection, all the data that is transferred amongst the vehicles within its coverage is collected by each RSU, removing their redundancy, exploiting algorithms of k-means for clustering the messages considering their characteristics, which include validity, sizes, and types, and lastly, communication parameter for each cluster is adjusted.

Reinforcement learning

Li, Wang, and Jiang (2017 pp.2217-2228) developed an online reinforcement learning method for balancing load to tackle the problem of user association in a dynamic setting for vehicles' networks. The achievement of the first relation is based on information on the recent framework by reinforcement learning. After a learning time, whereby the information of relations is gathered at the base location, the new relations outcomes will be adaptively and unswervingly gotten by a pattern of historical

association. Besides, Xu et al. (2014) used reinforcement learning for the design of vertical handoff for varied networks in vehicles. The fuzzy Q-learning method determines the connectivity of network using four kinds of information: strength value of receiving a signal speed of the vehicle, the quantity of data, and those users related with the network being targeted. In the learning-based tactic, users get linked to the most excellent network with no preceding knowledge on handoff conduct.

Deep reinforcement learning

Deep reinforcement learning facilitates the combination of artificial neural networks and architecture of reinforcement learning to enable the software-defined agents to acquire the best possible actions in the virtual setting to accomplish goals (Doddalinganavar, Tergundi, and Patil, 2019 pp. 81-86). Thus, it facilitates the unity of approximation of function and optimization of the target; the state-action pairs are mapped to anticipated rewards. Most algorithms for resource distribution for D2D-based networks are performed in a consolidated manner whereby information is collected by the fundamental controller and assessments for all vehicles to solve the optimization challenges (Li, Xu, Soong, and Ma, 2013 pp.15026-15047). Nonetheless, to attain information on the global network, the huge overhead will be incurred by the centralized control systems that dramatically grows with the vehicular network sizes. Therefore, a profound reinforcement learning basing on the mechanism for centralized resource allocation used in vehicular networks is applied where approximation with deep neural networks can be made by partial observations mapping of every vehicle agent with an optimized allocation of resources.

Deep learning

Deep learning is a kind of machine learning dealing with algorithms utilizing the brain function and structure known as artificial neural networks. Deep learning trains the computers to mimic what happens naturally to humans: learning by example. Deep learning is the vita technology applied in driverless cars, enabling them to identify a stop sign or to differentiate a lamppost from a pedestrian. Kang and Kang, (2016 pp. 1-5) suggested a system for intrusion detection for networks of vehicles basing on deep neural networks. The initialization of parameters by the unsupervised networks of deep belief as a stage of preprocessing. Then, deep neural networks' training buys packet data of high dimensions to determine the fundamental statistical properties packets that are normal and hacking and remove the corresponding characteristics.

Deep learning points to a learning technique of a machine by use of an architecture encompassing a multitude of ranked non-linear layers of processing steps. The depth of architecture refers to the

quantity of levels of the configuration of non-linear procedures learned in a function. Its design could be characterized into different types, which are, a generative and discriminative profound design, subject of the manner they are utilized (Ye et.al. 2018 pp.94-101). Deep discriminative structural design offers abilities for pattern classification including learning supervision like in the feed-forward conventional (ANN) artificial neural networks. Deep structure, that is, (DNN) deep neural network could be amplified using numerous concealed ANN arrangement layers.

Though, augmented neural networks are taught inadequately by learning of back-propagation with optimization of the gradient succession because of the challenge of disappearing gradient. In backpropagation, the error gradient surface is calculated in each stratum whereas the gradient descent is witnessed at abundant levels with the number of layers, therefore producing a tremendously sluggish merging speed (Liang, Ye, and Li, 2018 pp.124-135). To avoid the setback, the depth of generative design the correlation has characterized by the data detected and the associated classes is used for reorganizing constraints of the design, so-called the discriminative pre-training scheme.

In interconnecting nodes, weight factors in neighboring layers are trained competently by using a top-down method known as restricted Boltzmann Machines (RBM) by making an allowance for the nodes. Once the pre-training happens, the gradient descent process is engaged for fine-tuning with the learning supervised the same as in the convectional feed-forward ANN. (DBN) The deep belief networks are used as a probabilistic generative model, where numerous single concealed stochastic layers are are utilized for proficiency in solving the problem of vanishing gradient by units on top-bottom observed layer of data.

Examples of Neural Network architectures

Feed-Forward Neural Networks

This is the popular class in the physical applications of neural networks. The input is the initial layer and the last layer is output. If there exists more than one hidden layer, we refer to them as “deep” neural networks. They calculate a sequence of changes that alter the likenesses among cases.

Recurrent Neural Networks (RNN)

This usually possesses channeled sequences in their connection graph. Which means it is possible to get back to the beginning sometimes by the succeeding arrows. They could possess dynamics that are

complicated and this could prove them extremely tough to the sequence (Ye et.al. 2018 pp.94-101). They are much more realistic biologically.

Symmetrically Connected Networks (SCN)

They are similar to networks that are recurring, but the connecting elements are proportioned (they weigh in each direction equally). Symmetric networks are much easier to evaluate compared to recurrent networks. What they are capable of is more controlled because they are subject to a energy function. Nets without hidden units which are proportionally linked are called “Hopfield Nets.” Networks with hidden units and connected proportionally are called “Boltzmann machines.”

Perceptrons

Perceptrons represents the earliest form of neural networks, which entails models of a single computational neuron. Frank Rosenblatt popularized them during the early 1960s. They seemed to possess an algorithm used in learning that proved very powerful and numerous grand claims were put forward for what they were able to do after learning. A book published by Minsky and Papers in 1969 called “Perceptrons” explored what they were capable of including their limitations. The reviewed limitations were subject to each model of the neural network is what many people assumed. However, the learning procedure of perceptron is still broadly in use today for tasks that possess many millions of features with enormous feature vectors.

Convolutional Neural Networks (CNN)

These are composed of neurons that possess biases and weights that can be learned (Ye et.al. 2018 pp.94-101). Every neuron makes a dot product then tracks it optionally with non-linearity after receiving some inputs. The whole network expresses a single differentiable function mark: to class marks at the other end from the raw image pixels on one end. They remain with a function loss (e.g. SVM/Softmax) on the other (fully-connected) level. Convolutional Neural Networks can be utilized for all work related to the recognition of objects from hand-written digits to 3D objects.

Long short term memory (LSTM)

LSTM signifies the architecture for the artificial regular neural network (RNN) applied in deep learning and can learn lasting dependencies. Taylor, Leblanc, and Japkowicz, (2016, pp. 130-139) engaged Long Short Term Memory (LSTM) for detecting attacks on linked vehicles. The LSTM sensor is capable of recognizing the synthesized abnormalities with high precision since it has learned to the prediction of the subsequent word originating from every vehicle. The architecture of RNN is specifically vital for network security administration in VANETs.

Chapter5: Machine learning approaches in VANETS

Due to advancements in machine learning, (ML), vehicles ought to have safety measures which identify interruptions and make predictions so as to anticipate dangers. ML strategies have generally been used for this purpose. Q-learning has been utilized in upgrading the confirmation procedure against ridiculing assaults, hostile to sticking transmission and malware recognition. SVM (Support Vector Machine) can be used to recognize interruption and satirizing assaults. K-NN (K-Nearest Neighbour) on the other hand can be applied in identification of system interruption and malware (Ftaimi, and Mazri, 2020 pp.1-8). Vehicular clients can utilize Naïve Bayes to differentiate interruptions from mischievous activities. K-means can be used to distinguish congested driving conditions.

Hidden Markov models

A hidden Markov model is a type of statistical model which explains the development of observable events based on their internal factors that are not observable directly. The observed event is referred to as a 'symbol' and the invisible factor in the observation is referred to as a 'state'. This model is vital in implementing the predictive routing of VANETS through the exploitation of regularity of behaviors moved by vehicle for increasing the performance of the transmission (So, Sharma, and Petit, 2018 pp. 564-571). Since the vehicle movements display a lot of repetition, such as visiting certain places regularly and the regular contacts experienced while undertaking daily activities location of the future vehicle is predictable by using past traces knowledge and the hidden Markov model.

Variable-order Markov model

Variable-order Markov model (VOM) models are a vital category of models which cover the well-recognized Markov chain models. The VOM model applies the escape mechanism when addressing the problem of zero frequency, and the tree structure is used when decreasing the memory amount required in N-order Markov Model, where conditional distributions are stored in a large matrix. The inferior space complexity lowers the method efficiency while the prediction of results is influenced by the problem of zero frequency. Patterns of mobility of vehicles are extracted using variable-order Markov model from actual trace data in urban environment of vehicle networks that is applied in prediction of likely trajectories of the vehicles moving and develop an effective soft routing protocol based on prediction.

Recursive least squares

The algorithm of recursive least squares (RLS) involves the recursive application of the familiar algorithm of least squares (LS), such that every data point is accounted for in modification of parameter previously estimated through linear correlation supposed to have done the observed system modeling. The technique enables the Least-squares dynamic application in a series of time assimilated in real-time. In LS, there might have several equations of correlation and a set of dependent variables. The algorithm of RLS utilization in predicting vehicle movement with little complexity and high accuracy of prediction in VANETs. Moreover, the RLS algorithm is applied in the prediction of large scale channels based on vehicle location information and facilitates novel scheduling plan development for disseminating cooperative data in VANETs.

Regression

This section entails a discussion involving two types of regression. One of the two types of relapse is Support Vector Regression, which is generally a comprehensive modification of the SVM (Support Vector Machine) so as to fic complications relative to regression. The SVM helps to discover the hyper plane that allows the isolation of the information while being the furthest device from any perception (Ftaimi, and Mazri, 2020 pp.1-8). The other type of relapse is the linear regression. It basically focuses on creating a link between a variable X and Y, while enabling the expected value of Y for a quantified value X.

Clustering

K-means: The goal is to cluster information into K-class in the case where the provided data is similar. In this process, initial elements of k is selected as the centers of k clusters. Consequently, the items are placed in the nearest cluster, then the focal point of each group is recalculated to establish their centers of gravity (Ftaimi, and Mazri, 2020 pp.1-8). The process ought to be repeated till the center of gravity is stable. Moreover, the clustering class of ML algorithms includes a Density Based Spatial Clustering of Applications with Noise (DBSCAN). The main objective of DBSCAN entails categorizing the uncategorized dataset on the basis of their density.

Reinforcement Learning

Reinforcement learning is a technique which lets the calculation gain from its own faults by interacting with its surroundings to figure out how to arrive at the most appropriate decision (Ftaimi, and Mazri, 2020 pp.1-8). It works in a way that entails enforcing a penalty, when a wrong choice is

made and rewarding in the instance of a right choice. In a bid to get an ever increasing number of rewards, the calculation will put forth a noble effort to ensure optimal decision making. In this case the surroundings are represented on a Markov decision model.

Q-Learning

Q-learning represents a type of reinforcement learning relative to the sans model. Also, Q-learning can be presented as a method for non-concurrent dynamic programming. Noteworthy, the model is recognized as an off-policy reinforcement learning algorithms that attempts to identify the most appropriate decision to make with regards to the current state. It is regarded as an off-policy because the Q-learning model learns from the deeds external to the current method (Ftaimi, and Mazri, 2020 pp.1-8). The Q model determines the longing for the ultimate reward when making a move in a given state. Q learning avails the ability to figure out how well to act in Markovian spaces to relative agents. An agent engages the earth in two distinct techniques. Firstly, it can utilize the Q-tables as the basis of viewpoint and infer all possible outcomes in a given state. Secondly, the agent may opt to act randomly when making decisions.

K-Nearest Neighbors

K-Nearest Neighbors (KNN): The objective of KNN algorithm is to categorize novel input. According to Ftaimi, and Mazri, (2020 pp.1-8) The algorithm assesses the training set to check on what is the nearest point to the input. The algorithm employs the Euclidean distance gold measure also known as the Hamming distance to assess the closest point to the input. The novel input gets a label relative to the point closest to it. The K-nearest neighbor an extremely straight forward clustering technique, with less or no prior information on information dispersion. KNN algorithms must save all training data in memory which is perfect for small issues. In some instances, the KNN calculations may give wrong results when separating issues with high dimensional space (more than 10-15 measurements). NAÏVE BAYES on the other hand depends on the Bayes hypothesis in which represents a highly likely hypothesis. Notable NAÏVE BAYES mainly depends on conditional probability.

Chapter6: Comparison of ML approaches in ML

Ref#	Proposed Method	Security Attacks	Simulation Tools Used	Detection Accuracy	ML Used?
1) 2015	IDSAODV	Jellyfish and Black hole attacks	NS-2	82.20%	No
2) 2020	ANN	Data falsification-based attacks	NGSIM	N/A	Yes
3) 2012	WCMP	DDoS attack	NS-2	N/A	No
4) 2019	ML Algorithms	Multiple Misbehavior	Not mentioned	N/A	Yes
5) 2020	ML Algorithms	Multiple attacks	Not mentioned	N/A	Yes
6) 2017	Neural Network	Multiple Misbehavior	NGSIM Dataset	99%	Yes
7) 2011	Behavioral Features	Multiple Misbehavior	NCTUns-5.0 with TP rate	92%	Yes
8) 2019	Reputation Based ML	Multiple Attacks	VeReMi Dataset	97%	Yes
9) 2018	Framework	Real-Time DDOS	SUMO, OMNET++, Veins	99%	No
10) 2019	Algorithm	Malicious node	NS-2	95%	No
11) 2020	Framework	Multiple Misbehavior	SUMO LuST (DUA static)	N/A	No

12) 2016	Deep Learning	CAN Bus Attack	ROC Curve with OCTANE	97%	Yes
13) 2013	Framework	Multiple Misbehavior	Not mentioned	N/A	No
14) 2019	Sequence and driving matrix	Sybil Attacks	SUMO, OMNET++, Veins	>90%	No
15) 2017	ORLA	Load Balancing problem	Not mentioned	N/A	Yes
16) 2018	Framework	Multiple misbehavior	Not mentioned	N/A	Yes
17) 2016	Framework	Multiple Attacks	Not mentioned	N/A	No
18) 2013	Framework	Multiple Attacks	Not Mentioned	N/A	No
19) 2011	Algorithm	Multiple misbehavior	MATLAB	N/A	Yes
20) 2011	Data-centric MDS	Multiple Misbehavior	Not mentioned	N/A	No
21) 2019	ML based attack detection system	Wormhole attack	NS3, SUMO	99%	Yes
22) 2018	SVM, KNN in ML	Location Spoofing	VeReMi Dataset	94%	Yes
23) 2016	Congestion algorithm	DDOS	NS-2 and SUMO	92%	Yes

24) 2016	Anomaly Detection using ML	CAN Bus Attack	Not mentioned	99%	Yes
25) 2018	Framework	Multiple misbehavior	Not mentioned	N/A	Yes

Future research direction

Future researches in VANETs should focus on studying the techniques of security privacy and road users' trust. Future researchers need to address the problems that arise with the management of the public due to the increased vehicles numbers on the road. For example, an increase in overhead communication and longer certificate revocation lists (CRL). A CRL, comprises of certificates of all revoked vehicles (Ruj et.al. 2011 pp.1-5). Moreover, studies need to be conducted to provide solutions for less usage of public-key certificates. Future researches should also address connected vehicle technology aggregation algorithms that are efficient, appropriate, and secure. The studies should explore strategies to curb privacy violations of road users. Furthermore, more research needs to be conducted to substantiate the effectiveness and validity of this structure in VANNETs.

More studies should be done to develop intrusion detection methodologies for identifying malicious vehicles in VANETs. Also, research on the detection of misbehavior frameworks is vital to help in sensing integrate signal strength, validation of information, and position verification. Moreover, future studies should improve the available solutions for all malicious vehicles' eviction and remove them from the message dissemination contribution. Subsequently, studies on future technologies need to be done for the guidance of incorporation of routing protocol that is efficient and secure, representing a vital aspect of VANETs.

Conclusion

In conclusion, VANETs represents an important infrastructure in the vehicle routing technology and thus, it should provide adequate security guidelines as measures for secure communication. The enhancement of efficient driving and traffic safety represents a substantial motive in the application of VANETs. Periodic messages between nodes denotes a crucial principle during the application of VANETs. However, inevitable situations are triggered by inaccurate node transmission. Every transmitted packet contains: identity, time and position in addition to messages of safety concerning the sender. Moreover, a node with a tendency to misbehave may tamper with a propagated packet information. Therefore, the present study explore various machine learning approaches that are employed in VANETs to help enhance user privacy and security. Ultimately, the present study explores approaches in ML help to achieve better prospects to solve major security challenges in VANETs.

Bibliography

1) Alheeti, K.M.A., Gruebler, A. and McDonald-Maier, K.D., 2015, September. An intrusion detection system against black hole attacks on the communication network of self-driving cars. In 2015 sixth international conference on emerging security technologies (EST) (pp. 86-91). IEEE.
2) Almalki, S.A. and Song, J., 2020. A Review on Data Falsification-Based attacks In Cooperative Intelligent Transportation Systems. International Journal of Computer Science and Security (IJCSS), 14(2), p.22.
3) Biswas, S., Mišić, J. and Mišić, V., 2012, December. DDoS attack on WAVE-enabled VANET through synchronization. In 2012 IEEE Global Communications Conference (GLOBECOM) (pp. 1079-1084). IEEE.
4) Doddalinganavar, S.S., Tergundi, P.V. and Patil, R.S., 2019, July. Survey on Deep Reinforcement Learning Protocol in VANET. In *2019 1st International Conference on Advances in Information Technology (ICAIT)* (pp. 81-86). IEEE
5) Ftaimi, S. and Mazri, T., 2020, March. A comparative study of Machine learning algorithms for VANET networks. In Proceedings of the 3rd International Conference on Networking, Information Systems & Security (pp. 1-8).
6) Ghaleb, F. A., Zainal, A., Rassam, M. A., & Mohammed, F. (2017, November). An effective misbehavior detection model using artificial neural network for vehicular ad hoc network applications. In *2017 IEEE Conference on Application, Information and Network Security (AINS)* (pp. 13-18). IEEE.
7) Grover, J., Prajapati, N.K., Laxmi, V. and Gaur, M.S., 2011, July. Machine learning approach for multiple misbehavior detection in VANET. In *International conference on advances in computing and communications* (pp. 644-653). Springer, Berlin, Heidelberg.
8) Gyawali, S., & Qian, Y. (2019, May). Misbehavior detection using machine learning in vehicular communication networks. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)* (pp. 1-6). IEEE.

9) Haydari, A. and Yilmaz, Y., 2018, November. Real-time detection and mitigation of ddos attacks in intelligent transportation systems. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 157-163). IEEE.

10) Jeevitha, R. and Bhuvaneswari, N.S., 2019, February. Malicious node detection in VANET Session Hijacking Attack. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-6). IEEE.

11) Kamel, J., Wolf, M., van Der Heijden, R.W., Kaiser, A., Urien, P. and Kargl, F., 2020, June. VeReMi Extension: A Dataset for Comparable Evaluation of Misbehavior Detection in VANETs. In IEEE International Conference on Communications (ICC).

12) Kang, M.J. and Kang, J.W., 2016, May. A novel intrusion detection method using deep neural network for in-vehicle network security. In *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)* (pp. 1-5). IEEE.

13) Li, L., Xu, Y., Soong, B.H. and Ma, L., 2013. A reinforcement sensor embedded vertical handoff controller for vehicular heterogeneous wireless networks. Sensors, 13(11), pp.15026-15047.

14) Li, W. and Zhang, D., 2019, June. RSSI Sequence and Vehicle Driving Matrix Based Sybil Nodes Detection in VANET. In *2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN)* (pp. 763-767). IEEE.

15) Li, Z., Wang, C. and Jiang, C.J., 2017. User association for load balancing in vehicular networks: An online reinforcement learning approach. *IEEE Transactions on Intelligent Transportation Systems*, *18*(8), pp.2217-2228.

16) Liang, L., Ye, H. and Li, G.Y., 2018. Toward intelligent vehicular networks: A machine learning framework. IEEE Internet of Things Journal, 6(1), pp.124-135.

17) Pathan, A.S.K. ed., 2016. Security of self-organizing networks: MANET, WSN, WMN, VANET. CRC press.

18) Raw, R.S., Kumar, M. and Singh, N., 2013. Security challenges, issues and their solutions for VANET. International journal of network security & its applications, 5(5), p.95.

19) Rivas, D.A., Barceló-Ordinas, J.M., Zapata, M.G. and Morillo-Pozo, J.D., 2011. Security on VANETs: Privacy, misbehaving nodes, false information and secure data aggregation. Journal of Network and Computer Applications, 34(6), pp.1942-1955.

20) Ruj, S., Cavenaghi, M.A., Huang, Z., Nayak, A. and Stojmenovic, I., 2011, September. On data-centric misbehavior detection in VANETs. In 2011 IEEE Vehicular Technology Conference (VTC Fall) (pp. 1-5). IEEE.

21) Singh, P.K., Gupta, R.R., Nandi, S.K. and Nandi, S., 2019, March. Machine learning based approach to detect wormhole attack in VANETs. In Workshops of the International Conference on Advanced Information Networking and Applications (pp. 651-661). Springer, Cham.

22) So, S., Sharma, P. and Petit, J., 2018, December. Integrating plausibility checks and machine learning for misbehavior detection in VANET. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 564-571). IEEE.

23) Taherkhani, N., and Pierre, S., 2016. Centralized and localized data congestion control strategy for vehicular ad hoc networks using a machine learning clustering algorithm. *IEEE Transactions on Intelligent Transportation Systems*, *17*(11).

24) Taylor, A., Leblanc, S. and Japkowicz, N., 2016, October. Anomaly detection in automobile control network data with long short-term memory networks. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 130-139). IEEE.

25) Ye, H., Liang, L., Li, G.Y., Kim, J., Lu, L. and Wu, M., 2018. Machine learning for vehicular networks: Recent advances and application examples. IEEE Vehicular Technology Magazine, 13(2), pp.94-101.