Expert Systems With Applications 50 (2016) 40–54

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

# CEAP: SVM-based intelligent detection model for clustered vehicular ad hoc networks

Omar Abdel Wahab[a], Azzam Mourad[a,*], Hadi Otrok[b,c], Jamal Bentahar[c]

[a] Department of Computer science and Mathematics, Lebanese American University, Beirut, Lebanon
[b] Department of Electrical & Computer Engineering, Khalifa University of Science, Technology & Research, Abu Dhabi, UAE
[c] Concordia Institute for Information Systems Engineering, Montreal, Canada

ARTICLE INFO

*Keywords:*
Vehicular ad hoc network
Intrusion detection
High mobility
Support vector machine (SVM)
Malicious node
Training set size reduction

ABSTRACT

The infrastructureless and decentralized nature of Vehicular Ad Hoc Network (VANET) makes it quite vulnerable to different types of malicious attacks. Detecting such attacks has attracted several contributions in the past few years. Nonetheless, the applicability of the current detection mechanisms in the deployed vehicular networks is hindered by two main challenges imposed by the special characteristics of VANETs. The first challenge is related to the highly mobile nature of vehicles that complicates the processes of monitoring, buffering, and analyzing observations on these vehicles as they are continuously moving and changing their locations. The second challenge is concerned with the limited resources of the vehicles especially in terms of storage space that restricts the vehicles' capacity of storing a huge amount of observations and applying complex detection mechanisms. To tackle these challenges, we propose a multi-decision intelligent detection model called CEAP that complies with the highly mobile nature of VANET with increased detection rate and minimal overhead. The basic idea is to launch cooperative monitoring between vehicles to build a training dataset that is analyzed by the Support Vector Machine (SVM) learning technique in online and incremental fashions to classify the smart vehicles either cooperative or malicious. To adapt the proposed model to the high mobility, we design it on top of the VANET QoS-OLSR protocol, which is a clustering protocol that maintains the stability of the clusters and prolongs the network's lifetime by considering the mobility metrics of vehicles during clusters formation. To reduce the overhead of the proposed detection model and make it feasible for the resource-constrained nodes, we reduce the size of the training dataset by (1) restricting the data collection, storage, and analysis to concern only a set of specialized nodes (i.e., Multi-Point Relays) that are responsible for forwarding packets on behalf of their clusters; and (2) migrating only few tuples (i.e., support vectors) from one detection iteration to another. We propose as well a propagation algorithm that disseminates only the final decisions (instead of the whole dataset) among clusters with the aim of reducing the overhead of either exchanging results between each set of vehicles or repeating the detection steps for the already detected malicious vehicles. Simulation results show that our model is able to increase the accuracy of detections, enhance the attack detection rate, decrease the false positive rate, and improve the packet delivery ratio in the presence of high mobility compared to the classical SVM-based, Dempster–Shafer-based, and averaging-based detection techniques.



## 1. Introduction

Road accidents constitute the main leading cause of death for young people aged between 15 and 29 and the eight leading cause of mortality in general according to the World Health Organization (WHO)[1]. The organization warned that in 2030, road death will probably become the fifth-leading reason of death if precautionary measures are not taken. Vehicular Ad Hoc Network (VANET) (Wahab, Otrok, & Mourad, 2013b) is a multi-agent wireless network that is designed mainly to solve the traffic problems by allowing the smart vehicles to communicate with each other, as well

* Corresponding author. Tel.: +96170343643, +961 1 786456x1200; fax: +961 1 867098.
*E-mail addresses:* omar.abdelwahab@lau.edu.lb (O.A. Wahab), azzam.mourad@lau.edu.lb, azzammourad@gmail.com (A. Mourad), Hadi.Otrok@kustar.ac.ae (H. Otrok), bentahar@ciise.concordia.ca (J. Bentahar).

http://dx.doi.org/10.1016/j.eswa.2015.12.006
0957-4174/

[1] Global status report on road safety 2013: Supporting a decade of action, published by World Health Organization.

as with the roadside infrastructure such as traffic lights. For example, VANET enables vehicles to exchange emergency alerts in order to avoid collisions. Nevertheless, the applications of VANET are not restricted to the safety applications but englobe as well marketing, multi-media, and infotainment services (Huang, Chen, Chen, & Wu, 2009). Therefore, providing a good level of Quality of Service (QoS) is essential in such networks in order to ensure timely and accurate message delivery. Moreover, the highly mobile topology of vehicles imposes to take into consideration the mobility metrics in VANET to maintain the stability of the network.

Vehicular Ad Hoc Network Quality of Service Optimized Link State Routing (VANET QoS-OLSR) (Wahab et al., 2013b) is a clustering protocol that considers a tradeoff between the QoS requirements and the high mobility metrics in VANET. The protocol is based on electing a set of optimal cluster-heads in terms of QoS and dividing the network into clusters. To this end, a QoS function composed of several combinations of both QoS-based (bandwidth, connectivity) and mobility-based (velocity, residual distance) metrics is defined. The idea is to form stable clusters without sacrificing the QoS requirements. This function is used to elect the cluster-heads whose advantage is to facilitate the management of the clusters (Cheng, Yang, & Cao, 2013). These heads are then responsible for selecting a set of specific vehicles charged of transmitting the network topology information through messages called *Topology Control (TC)* and forwarding the packets. Such nodes are called *MultiPoint Relay (MPR)* nodes. VANET QoS-OLSR uses an algorithm based on Ant Colony Optimization (ACO) (Dorigo, Caro, & Gambardella, 1999) to select the MPRs satisfying the optimal path constraints. This algorithm takes into consideration the QoS function and End-to-End delay for this purpose. However, the problem arises when these selected intelligent MPRs behave maliciously and begin launching several attacks for the purpose of disrupting the network. Therefore, we propose in this paper an intelligent detection mechanism based on Support Vector Machine (SVM) learning technique to classify the vehicles in the clustered Vehicular Ad Hoc Networks either cooperative or malicious, while considering VANET QoS-OLSR as a starting point. The reason behind considering VANET QoS-OLSR comes from the fact that this protocol considers the formation of stable and long-living clusters, which is necessary in VANETs for any monitoring mechanism that requires buffering and comparing messages. As a case study, the packet dropping attack in which malicious MPRs drop the packets supposed to be retransmitted is considered. Such a misbehavior degrades the performance and lifetime of the network by isolating some cluster-heads. These cluster-heads will no longer receive the *TC* messages and hence will not be able to communicate with the other heads, which leads to a disconnected network. Although the packet dropping attack is considered as a case study in the subsequent sections, our model is generic and can be adapted to detect different types of malicious behaviors (e.g., Identity spoofing, Wormhole, etc.) by modifying the attributes used to build the classifiers accordingly.

The existing approaches that tackle the problem of malicious nodes in the domain of networks can be divided into two parts: detection-oriented approaches whose main goal is to identify the malicious nodes, and reaction-oriented approaches whose main goal is to deal with nodes after detection. This paper addresses the challenging problem of detecting and identifying the malicious vehicles in VANET, which is still an open research problem because of the challenges that are imposed by the special characteristics of VANET on any proposed detection mechanism. As for the reaction part, we have already proposed in our previous work (Wahab, Otrok, & Mourad, 2013a) a modified Tit-for-Tat strategy that could be adopted on top of our proposed detection model to control the relationships between vehicles after detection. The proposed strategy regulates the cooperation between vehicles in VANET after detection by propagating the detection results and advising the network nodes to fulfill the requests incoming from the detected cooperative vehicles and to drop those incoming from detected misbehaving vehicles.

Numerous detection models have been proposed in the literature for detecting misbehaving nodes in VANET. Nonetheless, the applicability of the existing models is hindered by the challenging characteristics of VANET. Specifically, the high mobility of vehicles complicates the process of monitoring, buffering, and analyzing observations as vehicles are continuously moving and changing their locations. Moreover, the limited resources of the vehicles especially in terms of storage space restricts the ability of vehicles to store and analyze the huge amount of observations that may be needed for accurate detections. To tackle these problems, we propose in this paper a cluster-based intelligent detection model for malicious vehicles called CEAP (Collection, Exchange, Analysis, and Propagation). The model combines the SVM classification technique (Han, Kamber, Pei, & Kaufmann, 2012) and watchdogs monitoring concept (Marti, Giuli, Lai, & Baker, 2000) in order to optimize the decision making process. The reasons behind choosing SVM as a classification technique are that (1) it is commonly known to be the best machine learning technique for binary (two classes) classification (Heller, Svore, Keromytis, & Stolfo, 2003; Hu, Liao, & Vemuri, 2003; Konar, Chakraborty, & Wang, 2005; Massimiliano, Alessandro, & Roi, 1997; Sung & Mukkamala, 2003; Vapnik, 1995); (2) it has been successfully used for intrusion detection (Heller et al., 2003; Hu et al., 2003; Sung & Mukkamala, 2003); (3) it is effective in high dimensional datasets with a large number of attributes (Shon & Moon, 2007); (4) unlike some other machine learning techniques such as Neural Networks (Haykin, 1998), SVM yields a unique solution since the optimality problem in SVM is convex (Auria & Moro, 2008); (5) it produces very accurate classifiers, is robust to noise, and minimizes the overfitting (Han et al., 2012).

In CEAP, the database is a result of a cooperative watchdogs monitoring process in which the watchdogs gather and share evidences about the behavior of the vehicles being classified. The SVM is then used by the cluster members in a distributed manner to distinguish well-behaving from misbehaving MPRs. For the sake of increasing the accuracy, we adapt SVM to work in both incremental and online fashion. This means that the training set is continuously growing by adding new training tuples (evidences) at each iteration and updated by what is learned from the previous iterations, which allows us to include additional training data without re-training from scratch (Laskov, Gehl, Krüger, & Müller, 2006). In order to mitigate the overhead and increase the efficiency of the model, CEAP exploits an important property of SVM, which states that only the support vectors (essential training tuples) are used to differentiate between classes. Thus, we consider that only these support vectors are kept from one iteration to another, which reduces the training set size in a considerable manner. In addition, the data is collected at the cluster-level targeting solely the MPR vehicles within each cluster, which are a set of specialized nodes responsible for packet forwarding. Thus, the database containing observations on the MPR vehicles exclusively is only communicated among the cluster members and only the final decisions are communicated among clusters.

**Contributions.** In summary, our contribution is a cluster-based lightweight intelligent detection model that uses the SVM machine learning technique in incremental and online fashion to classify the smart vehicles in multi-agent VANETs either cooperative or malicious. This model is able to:

- Increase the accuracy of detections, reduce the false alarms, and improve the routing process by cooperatively collecting