Investigation of the effectiveness of metric classification methods in identifying attacks in VANET

D Parfenov, I Bolodurina, L Grishina and A Zhigalov
Department of Applied Mathematics, Orenburg State University, 13 Prospekt Pobedy, Orenburg, 460018, Russia

E-mail: parfenovdi@mail.ru

Abstract. This paper discusses the problem of improving the efficiency of metric machine learning methods of identification attacks in vehicular adhoc networks (VANETs). The main idea of this research is to select the type of nonlinear functions for calculating the distances between the objects of the sample, describing the traffic of VANET using metric methods, such as the method of k-nearest neighbour with linearly decreasing weights and the Parzen window method. The analysis of the effectiveness of the methods considered was carried out on a synthetically generated sample with three different types of attacks on the network. Computational experiments have shown that the k-nearest neighbour method with decreasing weights based on an exponential function with base $a < 1$ is more efficient than the Parzen window method by about 0.3% and has an accuracy of 84.15%.

1. Introduction

With the development of information technologies, new forms of interaction between mobile devices and devices equipped with tools for interacting with intelligent systems have appeared. In this regard, the most relevant area of research at present is Vehicular Adhoc Networks (VANETs), which are designed to ensure road safety. Attacks on such networks cause critical and often emergency situations. Thus, the creation of the most effective protection mechanisms for wireless VANETs [1] is an urgent task at the present time.

The problem of identifying various types of attacks among benign network traffic is referred to as a classical machine learning problem – a classification problem. Intelligent methods of data analysis have proven their effectiveness in solving such problems [2]. However, with the constantly growing volume of incoming traffic, it becomes more and more difficult to identify attacks accurately enough and most importantly in a timely manner. In this regard, this study aims to improve the effectiveness of metric machine learning methods in identifying attacks on VANET. The main idea of this work is to select the type of nonlinear functions for calculating the distances between the objects of the sample, describing the traffic of VANET network using metric methods, such as the method of k-nearest neighbour (KNN) with linearly decreasing weights and the Parzen window method.

The rest of the paper is organized as follows. The second chapter provides an overview of existing literature sources on the topic under study. The third chapter presents the mathematical formulation of the problem of classifying attacks in VANET. The fourth chapter describes the application of metric machine learning methods. The fifth chapter contains experimental studies of the effectiveness of
various machine learning methods with various types of nonlinear functions for calculating the distances between sample objects, as well as the results obtained. The sixth chapter contains a conclusion.

2. Related works

Currently, studies related to analysis of attacks in VANETs cover most of the machine learning and other approaches to traffic analysis at various security layers. The key factors in assessing the performance of security algorithms are accuracy and detection speed. Let’s consider a number of scientific papers, which offer approaches to solving the problem of identifying attacks in VANETs, proposed by various authors.

The authors of study [3] considered cryptographic algorithms and intrusion detection systems based on the analysis of network characteristics as the main security tools. Comparison and assessment of security mechanisms has shown that the evolutionary growth of attacks cannot be contained by methods based on past developments, and timely modification of the available means of protection is required. The study [4] presents a Markov decision-making process with integration of machine learning and deep reinforcement learning methods while unloading the process of solving various task flows of the Internet of Vehicles management cycle. The most promising direction for the development of this system is the development of authentication and security mechanisms based on edge computing. The authors of study [5] noted the absence of a real dataset of VANET traffic that would cover most of the existing attacks on the network. In this regard, most of the research is conducted on synthetic data. In addition, the analysis of existing intrusion detection systems (IDS) showed that machine learning has not found wide application for the detection and prevention of distributed denial-of-service (DDoS) attacks in VANET and is the basis for the next generations of IDS. The study [6] proposes a machine/deep learning trust model approach for VANET security. A distinctive feature of the developed model is the modeling of trust as a process of classification and extraction of characteristics using a hybrid model of a Bayesian neural network. This approach links deep learning with probabilistic modeling to intelligently solve and efficiently generalize when calculating the trust level of the nodes in the network. The study [7] focuses on the application of an ensemble-based machine learning approach to classify abnormal vehicle behavior in VANET. As a result of a comparative analysis of the ensemble using five different methods, including Naive Bayes, AdaBoost, and basic classifiers, the authors confirmed the effectiveness of the developed attacks detection model. The authors of study [8] examined the application of machine learning methods for clustering and classifying intrusions in VANET using KNN and support vector machine (SVM) algorithms. The developed approach allows detecting DoS and fuzzy attacks that occur on the CAN bus. Experimental results showed that the algorithms have the same accuracy in identifying attacks, but KNN showed the best performance. A comparative analysis of these algorithms also presented in study [9], but within the framework of overcoming the imitation attack in VANET. In this case, it was experimentally established that the KNN algorithm has a 5% higher accuracy in identifying imitation attacks compared to SVM. In the study [10], the problem of detecting and classifying misbehavior related to location spoofing using the VeReMi dataset is considered. The research results showed that with the use of the KNN and SVM algorithms, the accuracy of the validation check detection increased by more than 20%, while keeping the recall within 5%.

The analysis of studies allows us to conclude the relevance of considered security problem in VANETs using data mining methods. This paper proposes improving the effectiveness of metric machine learning methods in identifying attacks in VANETs. The main idea of this research is to select the type of nonlinear functions for calculating the distances between the objects of the sample, describing the traffic of VANET, using metric methods, such as the method of k-nearest neighbors with linearly decreasing weights and the Parzen window method.

3. Statement of the problem of attacks identification in VANET

Let’s consider the incoming VANET traffic, which includes statistical information about the number and volume of transmitted data packets, the average transmission rate, and other characteristics of the
network flows. Within the framework of the formal classification problem statement, the received data on the traffic of the "objects-features" format form a set $X$.

The problem of building a model for attacks detection in VANET is in building a mapping $a : X \rightarrow Y$ – an algorithm, which determines the class of attack from the set $Y$ for each object from the set $X$.

The choice of a model and hyperparameters tuning is determined by the minimum additive error on the testing data:

$$Q(a, X') = \frac{1}{l} \sum_{l=1}^{L} \Lambda(a, x_l) \rightarrow \min,$$

(1)

where $\Lambda(a, x)$ is the loss function.

In this way, the efficiency of attack detection directly depends on the optimal choice of the machine learning method. Consequently, the selection of the type of nonlinear functions for calculating the distances between the sample objects is one of the most important tasks in the construction and training of models of metric algorithms, since on this phase the basis of the models under consideration is formed.

We will conduct a research of machine learning methods for identifying attacks in VANET using metric methods, such as the k-nearest neighbors method with linearly decreasing weights and the Parzen window method.

4. Metric classification methods

VANET security research has shown that machine learning plays an important role in identifying attacks and achieves acceptable recognition accuracy and speed. However, when analyzing a large traffic flow, it becomes more and more difficult to identify new types of attacks; the existing classification algorithms require revision and modification. tab

In this research, the issues of application of the most common machine learning algorithms for identifying attacks in VANET, as well as the selection of the most effective structure of the learning model, are considered.

4.1. Weighted k-nearest neighbour method

The main idea of this algorithm is to assign the classified object $u$ to the class to which the $k$ nearest training objects belong. The weighted $k$ nearest neighbors method differs from the basic KNN algorithm by introducing a strictly decreasing sequence of weights $w_i$ that specify the contribution of the $i$-th neighbor to the classification:

$$w(i, u) = [i \leq k]w_i; \quad a(u, X', k) = \arg \max_{y \in Y} \sum_{i=1}^{k} [Y^{(i)} = y]w_i.$$  

(2)

Note that the choice of the sequence $w_i$ has a significant impact on the classification accuracy and requires additional research to effectively solve various applied problems. The main idea in this case is to associate the weights of neighbors with the distance to them and their ordinal number. In addition, various types of dependencies between them are considered and presented in table 1.

**Table 1. Functions for decreasing weight sequences.**

| Name                        | Type of functional dependency |
|-----------------------------|-------------------------------|
| Inverse linear function    | $w_i = \frac{k - i}{k \cdot d_i}$; |
| Inverse quadratic function | $w_i = \frac{k - i}{k \cdot d_i^2}$; |
| Exponential function       | $w_i = \frac{k - i}{k \cdot 0.86^{d_i}}$; |
| with base $a < 1$           | $w_i = \frac{k - i}{k \cdot 0.86^{d_i}}$; |
Logarithmic function with base $a < 1$

$$w_i = \log_{0.9} d_i \cdot 10^{-19};$$

Arctang

$$w_i = \arctg \frac{k - i}{d_i}.$$ 

4.2. **Parzen window method**

The Parzen window method defines weights $w_i$ as a function of distance $\rho(u, x_i^{(i)})$, not the rank of $i$-th neighbor. Another distinctive feature of the algorithm is the introduction of a non-increasing kernel function $K(z)$, setting the $w(i, u) = K(\frac{1}{h} \rho(u, x_i^{(i)}))$:

$$a(u, X^j, k) = \arg \max_{y \in Y} \sum_{i=1}^{k} [y_{u}^{(i)} = y]K\left(\frac{\rho(u, x_i^{(i)})}{h}\right)$$

where $h$ is the width of the window, i.e. spherical neighborhood of the object $u$.

The main idea behind modifying this algorithm is to vary the kernel function $K(z)$. The types of kernels considered in the study are presented in table 2.

| Name                | Type of functional dependency |
|---------------------|-------------------------------|
| Epanechnikov kernel | $w_i = \frac{3}{4} \left(1 - \left(\frac{d_i}{h}\right)^2 \right) \left(\frac{d_i}{h}\right) \leq 1$; |
| Quadratic kernel    | $w_i = \frac{15}{16} \left(1 - \left(\frac{d_i}{h}\right)^2 \right) \left(\frac{d_i}{h}\right) \leq 1$; |
| Triangular kernel   | $w_i = \left(1 - \frac{d_i}{h}\right) \left(\frac{d_i}{h}\right) \leq 1$; |
| Gaussian kernel     | $w_i = (2\pi)^{-0.5} e^{-0.5 \left(\frac{x_i^{(i)}}{h}\right)^2}$; |

5. **Experiment**

Let’s research the classification accuracy of the considered machine learning algorithms with various types of nonlinear weight dependences $w_i$. The synthetically generated sample of VANET segment contains 32404 rows, including 7831 records of benign traffic, 8174 records of DDoS attack, 8167 records of intensive DDoS attack and 8232 records of UDP Flood attack.

5.1. **Weighted k-nearest neighbour method**

The results of attacks detection in VANET by the KNN method with various types of nonlinear weight dependences $w_i$ (see table 3) showed that the representation of a descending sequence using an exponential function with a base $a < 1$ allows achieving the most accurate classification ~ 84.15%. However, other types of functional dependencies on average showed 83.89% accuracy in detecting attacks.
Logarithmic function with base 
a <1
Arctang

A more detailed assessment of the effectiveness of the KNN method with the calculation of precision, recall, f1-score and support metrics is presented in table 4.

**Table 4.** Results of attacks detection in VANET with KNN.

|   | precision | recall | f1-score | support |
|---|-----------|--------|---------|---------|
| 0 | 0.8285    | 0.6703 | 0.7410  | 1953    |
| 1 | 0.8381    | 0.8882 | 0.8624  | 2075    |
| 2 | 0.8292    | 0.9048 | 0.8653  | 1996    |
| 3 | 0.8671    | 0.8950 | 0.8808  | 2077    |
| accuracy | -        | -      | 0.8415  | 8101    |
| macro avg | 0.8407  | 0.8396 | 0.8374  | 8101    |
| weighted avg | 0.8410 | 0.8415 | 0.8386  | 8101    |

5.2. Parzen window method
The results of attacks detection in VANET by the Parzen window method with different types of kernels (see table 5) showed that the use of the Gaussian kernel allows achieving the most accurate classification – 83.87%. The rest of the kernel types showed an average accuracy of 78.28% in detecting attacks, which is 5.5% lower than the best result.

**Table 5.** VANET traffic classification results based on the Parzen window method modification.

| Type of \( K(z) \) | \( h \) | Accuracy |
|---------------------|--------|----------|
| Epanechnikov kernel | \( 1.95 \times 10^1 \) | 0.7782    |
| Quadratic kernel    | \( 2.3 \times 10^1 \) | 0.7879    |
| Triangular kernel   | \( 2 \times 10^0 \) | 0.7825    |
| Gaussian kernel     | \( 2 \times 10^0 \) | 0.8387    |

A more detailed assessment of the effectiveness of the Parzen window method with calculating the precision, recall, f1-score and support metrics is presented in table 6.

**Table 6.** Results of attacks detection in VANET with the Parzen window method.

|   | precision | recall | f1-score | support |
|---|-----------|--------|---------|---------|
| 0 | 0.8048    | 0.6820 | 0.7384  | 1953    |
| 1 | 0.8416    | 0.8786 | 0.8597  | 2075    |
| 2 | 0.8281    | 0.9003 | 0.8627  | 1996    |
| 3 | 0.8730    | 0.8869 | 0.8799  | 2077    |
| accuracy | -        | -      | 0.8387  | 8101    |
| macro avg | 0.8369  | 0.8369 | 0.8352  | 8101    |
| weighted avg | 0.8375 | 0.8387 | 0.8364  | 8101    |
In this way, the results of the research of the accuracy of attacks detection in VANET showed that the KNN method with decreasing weights based on the exponential function with base $a < 1$ is more effective than the Parzen window method by about 0.3% and has an accuracy of 84.15%.

6. Conclusion

In this study, the problem of improving the efficiency of metric machine learning methods in attacks detection in VANET is considered. The main idea of this research is to select the type of nonlinear functions for calculating the distances between the objects of the sample, describing the traffic of VANET using metric methods, such as the method of $k$-nearest neighbors with linearly decreasing weights and the Parzen window method. The analysis of the effectiveness of the presented methods was carried out on a synthetically generated sample with 3 different types of attacks on the network.

Computational experiments have shown that when identifying attacks in VANET by the KNN method with a decreasing sequence $w_i$ based on an exponential function with base $a < 1$, the classification accuracy is 84.15%. Other types of functional dependencies on average showed 83.89% accuracy in detecting attacks. On the other hand, identifying attacks by the Parzen window method with a Gaussian kernel, the classification accuracy is 83.87%, which is 5.5% higher than the accuracy of the method with other types of kernels. In this way, the KNN method with decreasing weights based on the exponential function with base $a < 1$ is more efficient than the Parzen window method by about 0.3%.

Acknowledgement

The research was carried out with the financial support of the RFBR (project 20-57-53019) and the grant of the President of the Russian Federation for state support of leading scientific schools of the Russian Federation (NSh-2502.2020.9), as well as scholarships of the President of the Russian Federation to young scientists and postgraduates (SP-3652.2021.5).

References

[1] Ghori M R, Zamli K Z, Quosthoni N, Hisyam M and Montaser M 2018 Vehicular ad-hoc network (VANET): Review IEEE Int. Conf. on Innovative Research and Development ICIRD (Bangkok: IEEE) 1-6
[2] Quevedo C H O O, Quevedo A M B C, Campos G A, Gomes R L, Celestino J and Serhrouchni A 2020 An Intelligent Mechanism for Sybil Attacks Detection in VANETs IEEE Int. Conf. on Communications ICC (Dublin: IEEE) 1-6
[3] Malhi A, Batra S and Pannu H 2019 Security of Vehicular Ad-hoc Networks: A Comprehensive Survey Computers & Security 89 1-30
[4] Ali E S, Hasan M K, Hassan R, Saeed R A, Hassan M B, Islam S, Nafi N S and Bevinakoppa S 2021 Machine Learning Technologies for Secure Vehicular Communication in Internet of Vehicles: Recent Advances and Applications Security and Communication Networks 1-23
[5] Alrehani A M and Al-Haidari F 2019 Machine Learning Techniques to Detect DDoS Attacks on VANET System: A Survey 2nd Int. Conf. on Computer Applications & Information Security ICCAIS (Riyadh: IEEE) 1-6
[6] Eziama E, Tepe K, Balador A, Nwizege K S and Jaimes L M S 2018 Malicious Node Detection in Vehicular Ad-Hoc Network Using Machine Learning and Deep Learning IEEE Globecom Workshops GC Wkshps GC Wkshps (Abu Dhabi: IEEE) 1-6
[7] Grover J, Laxmi V and Gaur M S 2012 Misbehavior Detection Based on Ensemble Learning in VANET Advanced Computing, Networking and Security ed. P S Thilagam (Riyadh: Springer) 602-11
[8] Alshammari A, Zohdy M A, Debnath D and Corser G 2018 Classification Approach for Intrusion Detection in Vehicle Systems Wireless Engineering and Technology 9 79-94
[9] Mrugnayana S S and Sandeep A T 2020 Identifying Impersonation Attack in VANET using KNN and SVM Approach Int. Journal of Future Generation Communication and Networking 13(3) 1266-74
[10] So S, Sharma P and Petit J 2018 Integrating Plausibility Checks and Machine Learning for Misbehavior Detection in VANET 17th IEEE Int. Conf. on Machine Learning and Applications (ICMLA) (Orlando: IEEE) 564-71