Machine learning Approach for detection unknow activity in vehicular adhoc network

Atul B.Kathole 1, Dr.Dinesh N.Chaudhari 2

Research Scholar, J.D.I.E.T, Yavatmal 1
HOD, Department of Computer Sci & Engineering, J.D.I.E.T, Yavatmal, India 2,
atul.kathole1910@gmail.com 1

Abstract:

An enabling technology for providing security and valuable information in modern transport systems but susceptible to a number of attacks, ranging from auditing passively to aggressive interference is Vehicular ad hoc network also termed as VANET. IDSs i.e. Intrusion Detection Systems are critical tools for risk reduction when suspicious activities are detected. Additionally, collaborations of VANET vehicles increases the accuracy in detection by communicating interactions among their nodes. Due to this, machine learning framework of distribution is effective and can be scaled and applied to build collaborative detection algorithms over VANETs. Concern about privacy is a fundamental obstacle to collaborative learning, because data is exchange between nodes. A node which is malicious can get information that is sensitive about nodes other than itself through the data which is observed. This transcript proposes for VANETs, a collaborative IDS that safeguards machine learning privacy. In the algorithm proposed, the alternating multiplier direction approach is used to a class of empirical risk minimization problems and an intrusion detection classifier is trained in the VANET. The use of privacy differential is done to capture the notation of privacy and apply a vector approach of dual disturbance to dynamically varying privacy.

Keyword: *VANET, Security, Algorithm, IDS.*

I. Introduction:

With an increase in vehicles on the road and autonomous vehicles advancing rapidly, road safety is becoming an increasing problem. VANET provides a communication mechanism for disseminating safety, traffic control, navigation and road service information. From eavesdropping passively to direct interference [1] VANETs are therefore considered vulnerable to a number of attacks. For example, to access similar tools, such as toll services, an intruder may search and replay other vehicle messages. An intruder can interfere with a targeted vehicle, its identity is impersonated and a false warning is sent that could interfere with road traffic [1].

Machine learning also termed as ML uses an artificial intelligence i.e. AI technique to teach a computer on things that are not known and to take successful decisions precisely. ML finds its usage in nearly every field, such as manufacturing, robotics, arts, biotechnology, smart automated transportation systems and automated systems, it has become widespread since it is available at less-cost and are very capable machines (that is the computing power is high and storage of data is massive), also the existence of large data volume. It allows decision making that is fast and intelligent to improve the performance of the system which includes efficient energy, quality of service (QoS) and reliability [1]. Owing to exponential population growth and automobile expansion, congestion of road and health which has become dynamic and vexing issues in a lot of metropolitan area. Every year about 1.25 million people die from road accidents worldwide, it is the main reason of death between people aged 15 to 29[2]. The congestion results in costly delays, heat, emissions and wasted fuel. In 2017 it cost $305 billion in congestion in the United States [3]. A smart and efficient transport network is capable of providing less road accidents, an eco-friendly environment and smooth traffic flow, in turn this increases performance. The VANET which stands for Vehicular Ad Hoc Network finds its application to enhance safety on the road and improvement of traffic congestion, especially during peak hours, in order to reduce travel time for travelers. The rapid increase in the demand of wireless devices has directed us to the requirement of a huge spectrum which accommodates large volume capacity allocation over demand and has been an obstacle to the implementation, accommodate, and next-generation switching technology scaling, this includes smart cities, video streaming services of high-definition 3D, augmented reality, Internet of Things also known as IoT and VR i.e. virtual reality.

The exponential growth of wireless devices has led to the need for a vast spectrum to support high-volume data allocation compared with the demand) has become a hindrance to the deployment, support, and scaling of next generation applications for commuters, including the Internet of Things (IoT), smart cities, virtual reality, augmented reality, and high-definition 3D video streaming services.

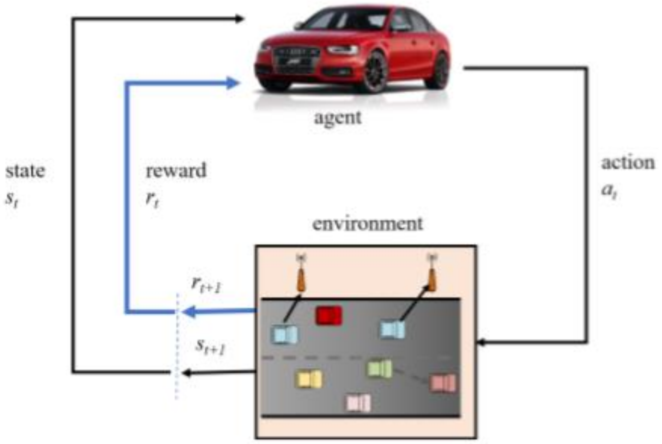

Figure 01: VANET architecture using intermediator

That's why this work proposes a shared VANET IDS focusing on privacy-conserving machine-learning. First, ADMM is used to create a distributed problem of minimizing empirical risk on a VANET, so that the training of the classifier can be done in a decentral manner to identify whether an operation is natural or an attack. At CIDS, to catch the notation of privacy in distributed machine learning, we are extending differential privacy to dynamic differential privacy, and proposing a conserving approach for privacy, which is the dual variable destruction. Also, we analyze the DVP 's usefulness and describe the DVP 's basic interaction between security and privacy by devising a convex optimization problem and running data sets based on statistical tests to illustrate the optimum existence of the privacy mechanisms.

The main contributions of this paper are summarized as follows:

1. We are introducing a machine-learning-centric platform for exchange of information and sharing of knowledge at VANETs.
2. Usage of ML to note the distributed essence of a VANET and create shared learning based on a VANET algorithm that is regularized.
3. We measure the potential performance that is determined by the least amount of training data that is required for a small error to occur.

II. Related Work

Many works have studied various architectures of intrusion detection systems that are well-suited to MANET [3].

Numerous studies have researched on the structures of different system of detection for intrusion that are well matched to MANET [3]. MANET's architecture are mostly divisible into 3 classes. distributed and cooperative IDS is the first one, catching the scattered nature of MANET which has the potential to extend through network cooperation. In [10], for example, Zhang and Lee used this form of MANET to construct a model for a global, cooperative IDS. And Albers et al. Using mobile agents in [11] Suggested a local IDS based shared IDS. The local IDS is implemented for local network-based security issues on each MANET network, and can be extended by continuing local IDS coordination over the MANET to address global security issues. The second type is a hierarchical IDS model which mixes relational and cooperative architectures. In [12] Sterne et al. used multilevel clustering to construct a complex hierarchical IDS. The third architecture makes use of the idea of a virtual agent capable of roaming around the huge network. In this system each agent that are mobile is allocated to operate on a given task; one or more mobile agents are then spread within the MANET to each node. Previous studies include the Kachirski and Guha in's work [13] who introduced dispersed IDSs using multiple mobile agent-based sensors; hence, the load of work is spread by segregating practical work and delegating roles to individual agents. In literature ML and data processing also researched. These strategies allow IDS to learn continuously about threats and their operations, develop security system expertise, relate unusual events and anticipate an attack. Researchers were studying unattended learning in IDSs, such as the clustering process, which is an unattended exploration of trends. There are many strategies for clustering data which is not labeled; example, a density-based spatial clustering of applications with noise clustering algorithms [14] was introduced by Blowers and Williams to merge normal network packets with anomalous ones. Many examples on clustering include the clustering based on hierarchy [15] and the K-means [16]. IDS literature also covers guided learning, for example support for vectors [17]. For example, Wagner et al. [18] implemented an SVM classifier of a single class, using a kernel of new window to locate the defect based on the data's time position. Many approaches are included in the supervised learning such as decision trees [19, 20, 21], artificial neural networks [22, 23] and sequential data aggregation [24, 25]. Works were also carried out on intrusion-prediction-dependent detection which used non-ML technique [26, 27]. Nidhal et al. [28], for example, developed a game-theoretical model for detecting intrusions in the VANET. This model forecasts a probable denial-of - service attack in the future against the controlled nodes. There are a host of studies on the application of differential privacy to machine learning in the field of privacy differentials [29, 30, 31, 32, 33]. For example, in probabilistic terms, Kasiviswanathan et al. [29] performed a generic method for the roughly learning correctly. Since researching differential privacy theory (e.g., [9, 34, 35]), a literature body has been exploring the interaction within privacy and performance in ML. A number of experts, too, are concentrating on the health of the transmitted differentials. In [36] implementing cryptographic protocols, Eigner and Maffei came up with the mechanism for verification of the distributed difference protection which is completely automated. In order to preserve the privacy of constraint collection, Han et al. [37] presented an algorithm which is differentially private to solve a distributed restricted optimization based on distributed simulated gradient descent. Hale et al. [38] made use of a cloud computer to conduct differential computations privately, so the transmission of outcomes to every agent through the network did not leak each agent's state. This paper uses distributed machine learning to create a mutual IDS to overcome the privacy barrier by incorporating the concept of dynamic differential privacy to maintain the identity of the research dataset used in the learning process.

ML implementations have been discussed in numerous fields, such as traffic modeling, filtering, and network classification. It submitted a Deep Learning Survey at [26]. In [27] Gosavi was presented the central theory of reinforcement learning methods, which is a technique of ML. A systematic analysis of the ML techniques was undertaken in CR [28]. Usage of AI has been given in various CR implementations in [29]. Many ML applications were debated in CRNs in [30]. Extensive information about the use of various techniques of AI in CRNs has been addressed [22]. A detailed overview of the use of various ML methods in dynamic spectrum access (DSA) is given [15]. Latest discoveries and uses of ML in VANETs have been addressed in [31]. Analysis in detail of the different ML methods used in VANETs was described in [21], and various applications for VANET AI were discussed in [32]. A short survey was performed on CR at

[33]. They discussed here the CR's basic principles and its other measures, taxonomies, difficulties and challenges. For comprehensive CR information please see [34],[35]. A description of the CR cycle was given in [36], which consisted of four stages of processes for a CR, that is spectrum sensing, perception, reasoning and adaptation (see Section II.C). Studies of various spectrum sensing techniques were performed at [37]– [40]. Knowledge on spectrum diversity and its concerns was addressed at [41]. A reform of spectrum control was carried out at [12].

The open loop strategies block interference from happening in the former case. However, congestion is regulated in closed loop approaches, after it has been identified [28]. identification of congestion can be achieved using measuring approaches that senses the no. of messages in queue, channel use level and occupancy of channel [29]. As described in the introduction, the congestion management strategies of VANET are categorized into hybrid strategies and strategies based on prices, resources, CSMA / CA, priority setting and scheduling [17]. Below we are going to discuss those tactics.

III. Proposed work:

This part of the research describes the proposed model architecture that involves several VANET blocks of building. VANET typically consists of application AU, on-board - RSU and OBU modules. Contact among the OBUs (car - car) or an OBU with an RSU (vehicle to infrastructure) is on the basis of the Vibration, i.e. wireless communication in the vehicle environment [3]. RSUs might connect also to other infrastructures, say for example other RSUs and control centers for traffic, and other wireless communication (infrastructure to infrastructure) is used to link with them. — It's got an OBU, plus any AU or so. It also has a collection of sensors that the OBU uses to collect information with other RSUs or OBUs and share information. Information on the three key parts to a VANET 's architecture that are presented for interested readers in Appendix A. A local PML-CIDS agent is present in each vehicle as seen in Figure 2, this tracks local operations that include those in the OBU and AU communications.

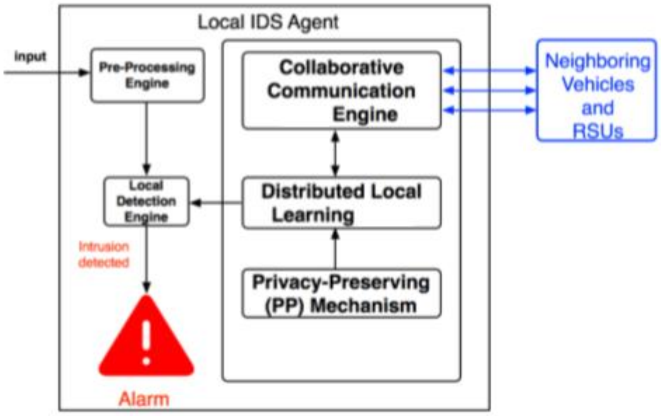
Local IDS Agent
input
Pre-Processing Engine
Collaborative Communication Engine
Neighboring Vehicles and RSUs
Local Detection Engine
Distributed Local Learning
Intrusion detected
Privacy-Preserving (PP) Mechanism
Alarm

Figure 2. Architecture of Proposed IDS System using ML

The collaborative framework is conceptually composed of three essential elements, a detection engine which is local to the system, a collaborative ML engine, and a pre-processing engine that preserves privacy. In Algorithm 1 logic flow is given. The collection and pre-processing of data from the VANET framework takes place in the pre-processing engine which determines the activities of the vehicle system in real time.

The pre-processed system Algorithm 1

```
Input: Real-time VANET system data: Local audit data ?ow and activity logs.
Step 1: The pre-processing engine collects and preprocesses the real-time VANET system data, by data transformation, features selection, and data normalization.
if the classi?er needs update then
    Step 2: The IDS engine is initiated and local training dataset is loaded. And updated classi?er is obtained.
    Step 3: The local detection engine uses the newly updated classi?er to analyze the real-time VANET system data.
    If any activities are classi?ed as intrusions, the local detection engine triggers the alarm.
else
    Step 2: The local detection engine uses the current classi?er to analyze the real-time VANET system data and triggers the alarm when any activities are classi? ed as intrusions.
end if
```

Mathematical Model

$$d(s,n,t,r,d)= \alpha1 \, . \, m(s,n,t,r,d) + \alpha2 \, . \, g(s,n,t,r,d)$$

where,

d: Resultant malicious node detection scheme,

m: Cooperative malicious node detection scheme,

g: Malicious node detection scheme using IDS technique,

$\alpha1$, $\alpha2$: Optimization factor that is constant between [0,1], such that $\alpha1 + \alpha2=1$.

The data is later processed using the detection system locally, using techniques like clustering. If the operator wants to change the internal classifier then the engine will run. The Local Detection Algorithm analyzes the device data using the newly acquired classifier. Another means the existing classifier will be included in the intrusion list. The warning is activated if any intruder is marked. An integral component of distributed local learning (DLL), privacy-preservation (PP) method, and collaborative communications engine (CC) will be further elaborated in Appendix B.

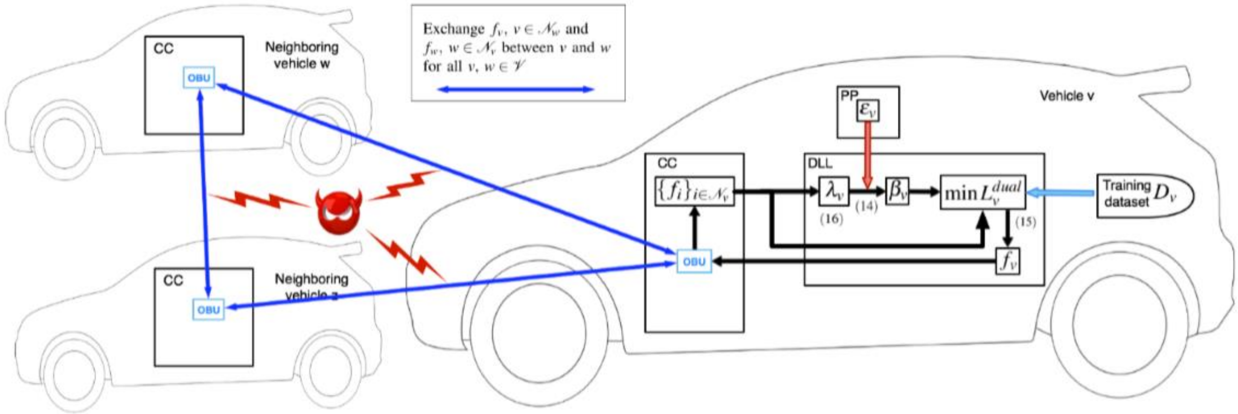

Figure 03: Working model of IDS system.

IV. Outcome:

Using this model for learning is a continual advancement and there is no definitive learning cycle for every single vehicle v. In fact, every single v makes decision is to when a new collaborative learning begins to update the last updated classifier f1 v or avoid collaborative learning and keep the recently updated classifier f2 v as the existing classifier for intrusion purposes. Continuous learning is required as the training data continue to grow. The machine learning algorithm will take advantage of regular data set updates to constantly analyze specific types of threats and their behavior to improve security system awareness.

V. Conclusion:

In this article, we established an architecture of system for collaborative intrusion detection using distributable ML for privacy protection. A scheme of privacy-preservation for centralized collaborative learning is essential for private collaboration maintenance; otherwise the program's scattered learning would itself lead to the leakage of the private instruction data. It implements a shared intrusion prevention program based on machine-learning for privacy-conservation. The alternating direction method of multipliers is used to remove the central problem of empirical risk minimization which designs collaborative learning via the hierarchical which is well suited to the design of the VANET systems.

In order to preserve the privacy of training data security, the implementation of dynamic differential privacy is done and creation of dual variable perturbation is done by manipulating the dual variable. We have technically evaluated the performance which is determined by the smallest amount of training data needed to train a low error classifier. Numeric studies have explored the tension between privacy and protection. The is the data set which is used during experiments. We have suggested a design theory that chooses the best value of the αv(t) data protection parameter by providing a solution to a problem of optimization that also optimizes privacy and security. The studies have explored the effect during joint learning of the different sizes of VANET, and VANET topology. As future research, we plan to explore both supervised virtual IDS and machine learning which is not supervised, and expand the dynamic of differential privacy to cover multiple ML techniques. We plan to learn rapid incremental learning techniques which can be used in regular IDS-based machine-learning updates

Reference:

1. L. Liang, H. Ye, and G. Y. Li, “Towards Intelligent Vehicular Networks: A Machine Learning Framework,” IEEE Internet Things J., vol. PP, no. c, p. 1, 2018.

2. J. Qadir, “Artificial intelligence based cognitive routing for cognitive radio networks,” Artif. Intell. Rev., vol. 45, no. 1, pp. 25–96, 2016.
3. B. Khalfi, A. Zaid, and B. Hamdaoui, “When machine learning meets compressive sampling for wideband spectrum sensing,” in 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), 2017, pp. 1120–1125.
4. C. Chembe, D. Kunda, I. Ahmedy, R. Md Noor, A. Q. Md Sabri, and M. A. Ngadi, “Infrastructure based spectrum sensing scheme in VANET using reinforcement learning,” Veh. Commun., vol. 18, p. 100161, 2019.
5. R. Boutaba et al., “A comprehensive survey on machine learning for networking: evolution, applications and research opportunities,” J. Internet Serv. Appl., vol. 9, no. 1, p. 16, 2018.
6. S. Pouyanfar et al., “A Survey on Deep Learning: Algorithms, Techniques, and Applications,” ACM Comput. Surv., vol. 51, no. 5, pp. 92:1--92:36, Sep. 2018.
7. Y. Gordienko et al., "Deep learning with lung segmentation and bone shadow exclusion techniques for chest x-ray analysis of lung cancer," in International Conference on Theory and Applications of Fuzzy Systems and Soft Computing, 2018, pp. 638-647: Springer.
8. M. T. Brown and J. K. Bussell, "Medication adherence: WHO cares?," Mayo Clinic proceedings, vol. 86, no. 4, pp. 304-314, 2011.
9. K. Teng. (2012, May). What Is Personalized Healthcare? From patients to medications, one size does not fit all. Available: https://health.clevelandclinic.org/what-is-personalizedhealthcare.
10. Ara and A. Ara, "Case study: Integrating IoT, streaming analytics and machine learning to improve intelligent diabetes management system," in 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017, pp. 3179-3182.
11. R. Vargheese and Y. Viniotis, "Influencing data availability in IoT enabled cloud based e-health in a 30 day readmission context," in 10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, 2014, pp. 475-480.
12. M. D. Naylor, L. H. Aiken, E. T. Kurtzman, D. M. Olds, and K. B. Hirschman, "The importance of transitional care in achieving health reform," Health affairs, vol. 30, no. 4, pp. 746-754, 2011.
13. S. Tilson and G. J. Hoffman, "Addressing Medicare hospital readmissions," Congressional Research Service, 2012.
14. S. Shahrestani, "Assistive IoT: Deployment Scenarios and Challenges," in Internet of Things and Smart Environments: Springer, 2017, pp. 75-95.
15. M. J. Rothman, S. I. Rothman, and J. Beals IV, "Development and validation of a continuous measure of patient condition using the Electronic Medical Record," Journal of biomedical informatics, vol. 46, no. 5, pp. 837-848, 2013.

16. H. Ye, L. Liang, G. Y. Li, J. Kim, L. Lu, and M. Wu, “Machine Learning for Vehicular Networks: Recent Advances and Application Examples,” IEEE Veh. Technol. Mag., vol. 13, no. 2, pp. 94–101, 2018.
17. W. Tong, A. Hussain, W. X. Bo, and S. Maharjan, “Artificial Intelligence for Vehicle-to-Everything: A Survey,” IEEE Access, vol. 7, pp. 10823–10843, 2019.
18. D. M. Alias and G.K. Ragesh, “Cognitive Radio Networks : A Survey,” in IEEE WiSPNET 2016, 2016, pp. 1981–1986.
19. X. Lu, P. Wang, D. Niyato, and E. Hossain, “Dynamic spectrum access in cognitive radio networks with RF energy harvesting,” IEEE Wirel. Commun., vol. 21, no. 3, pp. 102–110, 2014.
20. G. Zheng, Z. Ho, and E. A. Jorswieck, “Information and Energy Cooperation in Cognitive Radio Networks,” IEEE Trans. Signal Process., vol. 62, no. 9, pp. 2290–2303, 2014.
21. Beibei Wang and K. J. R. Liu, “Advances in cognitive radio networks: A survey,” IEEE J. Sel. Top. Signal Process., vol. 5, no. 1, pp. 5–23, 2011.

22. S. Pandit and G. Singh, “Spectrum Sensing in Cognitive Radio Networks: Potential Challenges and Future Perspective BT - Spectrum Sharing in Cognitive Radio Networks: Medium Access Control Protocol Based Approach,” S. Pandit and G. Singh, Eds. Cham: Springer International Publishing, 2017, pp. 35–75.
23. N. Muchandi and R. Khanai, “Cognitive Radio Spectrum Sensing : A Survey,” Int. Conf. Electr. Electron. Optim. Tech. - 2016 Cogn., pp. 3233–3237, 2016.
24. S. M. Baby and M. James, “A Comparative Study on Various Spectrum Sharing Techniques,” Procedia Technol., vol. 25, no. Raerest, pp. 613–620, 2016.
25. D. Sun, T. Song, B. Gu, X. Li, J. Hu, and M. Liu, “Spectrum Sensing and the Utilization of Spectrum Opportunity Tradeoff in