

TON-IoT: Detection of Attacks on Internet of Things in Vehicular Networks

1st Anshika Sharma

*Chitkara University Institute of
Engineering and Technology,
Chitkara University,
Rajpura, 140401, Punjab, India
anshika.sharma@chitkara.edu.in*

2nd Himanshi Babbar

*Chitkara University Institute of
Engineering and Technology,
Chitkara University,
Rajpura, 140401, Punjab, India
himanshi.babbar@chitkara.edu.in*

3rd Avinash Sharma*

*(Corresponding) Department of
computer Science and Engineering, MM
Engineering College, Maharishi
Markandeshwar (Deemed to be
University), Mullana-
Ambala, Haryana, India
asharma@mmumullana.org*

Abstract—During the recent era, due to the exponential rise in the Internet of things (IoT) enabled devices around the globe, IoT and Machine learning (ML) has come out as usable and efficient approach to supply productive solutions in this environment. Vehicular networks (VN) are considered to be the most important application domain where ML-based techniques are generated to address common attack issues and present insightful discussions. It is the union of smart transport and internet systems which is responsible for the passenger's safety and security as the attack threats have been growing very rapidly in VN. The main objective is to overwhelm the targeted IoT devices with malicious data traffic in VN. To solve the above-mentioned issues, this paper covers numerous types of attacks (backdoor, injection, Distributed Denial of service (DDoS), ransomware, password, scanning, Man in the middle (MITM) and cross-site scripting (XSS)) on VN deploying ML techniques using intrusion detection system (IDS) in the IoT based on the TON-IoT dataset. Due to the fact that it contains a variety of normal and malicious activities for various IoT services as well as heterogeneous data suppliers, TON-IoT provides a number of advantages that are completely missing from state-of-the-art datasets. The performance metrics are evaluated for the attack detection namely accuracy, precision, recall and F1-score using the three ML methods Random Forest (RF), Naive Bayes (NB) and K-Nearest Neighbour (KNN). Finally, the results conclude that amongst the three ML methods, KNN gives the highest accuracy rate i.e. 98.2%, while RF and NB give a 94% and 70% accuracy rate utilizing the ToN-IoT dataset.

Index Terms—Machine Learning, Vehicular Network, Security, Intrusion Detection System, ToN-IoT dataset.

I. INTRODUCTION

According to the report of global status given by the World Health Organisation (WHO) on road safety, road accidents are the major cause of people's demise within the age group 6-30 years [1]. A VN is used to manage multiple applications in transportation systems such as road safety, vehicle speed and traffic management etc. These networks are designed to maintain secure and efficient transportation through secure communication networks [2]. A VN can communicate through wired and wireless networks and the data is transferred between vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), infrastructure-to-infrastructure (I2I), etc. However, these networks throw some security-related issues that need to be examined while creating vehicular security and privacy solutions [3].

Although IoT increases effectiveness through management, the

risk of attack threats is also increasing due to the deficiency of security measures. VN now face various challenges, including a lack of stability, adaptability, and modifiability when attempting to apply the services in a large-scale context. Due to VN system dynamics and lack of effective routing, managing the network and traffic in these systems is difficult. This results in network congestion and difficulty with network throughput [4].

Recently, ML has become a powerful tool for accelerating security and improving attack prediction accuracy [5]. A broad range of techniques is used in this paper to overcome the different security-related issues. IDSs are used as protection for the server to detect attack threats that break security boundaries in VN. To accurately and effectively assess the effectiveness of IoT security solutions, it is essential to estimate IDS methodologies and use the TON-IoT dataset, which depicts realistic IoT applications [6].

The main goal of this paper is to present a recent and effective ToN-IoT dataset that can be used to analyze security solutions using intrusion detection methods in VN. The proposed dataset is available publicly for the use of research-related communities [7]. However, the features and description of the TON-IoT dataset are also provided here.

The main contribution of the presented paper is expressed as:

- An architecture of VN using IDS is proposed to solve a wide range of security challenges including attack threats, privacy and intrusion detection etc. and refine the communications from V2V and V2I to protect the network from the attack threats.

- A recent TON-IoT dataset is analyzed that contains a total of 46 features. After the data preprocessing, nearly 14 features are taken out from the total features for analysis purposes.

- The performance evaluation of three ML methods is also provided on the proposed dataset for the detection of attack threats.

This paper is summarised as follows: Section II presents the related work in the field of existing datasets, ML methods and the IDSs in VN. Section III provides an architecture of VN using the IDS. A summary of the TON-IoT dataset and the data preprocessing steps are given in Section IV. The ML methods are also discussed in this section. An evaluation and the result analysis of the performance metrics are shown in

Section V. Finally, Section VI concludes the result of the presented paper.

II. RELATED WORK

This section presents some information and related research work about IDSs based on machine learning algorithms. Numerous datasets are also discussed for the evaluation purpose of IDS in VN. A brief overview related to this paper has been provided here with the different approaches. The existing literature of attack detection in VN is also presented in Table I.

Basavaraj et al. [2] proposed a framework for intrusion detection in VN. He also presents the solution for the attack mitigation on the Controller Area Network (CAN) bus protocol with the help of IDS based on the Deep Neural Network (DNN).

X.Li et al. [13] use classification techniques and offer update strategies for the models depending on how well the VN cloud can promptly give a modest amount of classified data for a new threat.

M. Booij et al. [6] examine the novel TON-IoT dataset containing 46 features. He presents the proposed dataset's architecture and statistical analysis and creates some useful observations into the features between training and testing datasets.

Alsaedi et al. [7] discussed the proposed TON-IoT dataset and its features based on the Industrial Internet of Things (IIoT). The authors measure some evaluation metrics such as accuracy, recall, F1-score and precision to evaluate the performance of various ML and DL approaches for intrusion detection purposes.

Nour et al. [14] designed a new testbed architecture of IoT for gathering federated data involving both normal and malicious traffic, to provide various IoT services such as Software-Defined Networks (SDN) and Network Function Virtualization (NFV).

Vivekanadam B. [15] provides a hybrid technique for identifying and combating HNN-based Denial of Sleep Attacks and mobile sink.

III. ARCHITECTURE OF VEHICULAR NETWORK USING IDS

The vehicular network is in danger due to various attack threats. An architecture proposed for the detection of attacks in a VN is shown in Fig. 1 which contains a cloud server, Base Station (BS), Road Side Units (RSU), Basic Safety Messages (BSM) and some smart vehicles [4]. This strategy proposes a secure architecture for VN to secure the data flow between V2V, I2I and V2I etc.

A comprehensive description of the proposed architecture is described below:

A. Road Side Units

A vehicular network allows smart vehicles to connect to RSUs. RSU is a computing device that collects data from a sensing region along a road and provides connectivity support to transmit it to the server.

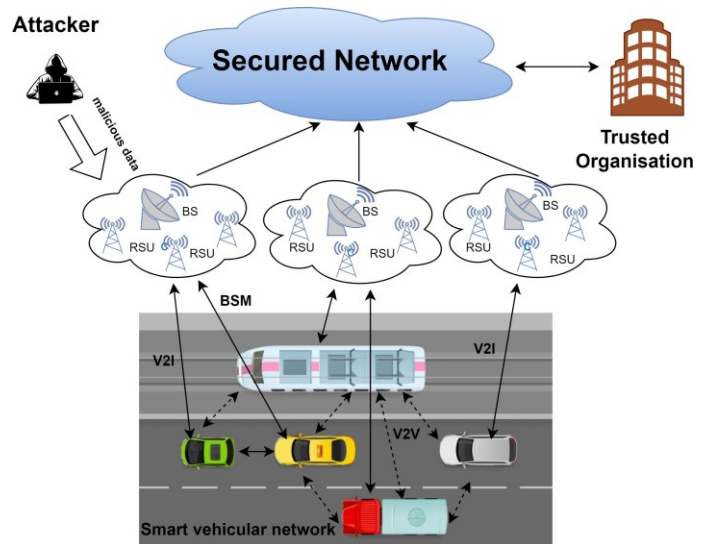


Figure 1: Proposed architecture for IDS based vehicular networks

B. Base Station

A base station is a transceiver or the main connection point responsible for the communication of wireless devices. It further links the devices to other devices/networks.

C. Nodes

In a VN, the nodes are the vehicles that are responsible for the communication between the vehicles and roadsides. It comprises V2V and V2I communications, where V2V can communicate the data about the vehicles' speed and location, and V2I is the communication model that exchanges data between vehicles and the road infrastructure. It gives information to drivers that they need to know for safety purposes.

D. Cloud server

Several computations such as driver registration, vehicle registrations, vehicle speed calculations and traffic situation etc. are performed at this layer. All the data is stored and analyzed in the database and managed through the cloud server.

E. Trusted Organisation

The responsibilities of the trusted organisation are explained below:

- Vehicles registrations
- To manage the security related issues

A secure and efficient architecture for the VN is designed in such a way that protects the vehicle's sensitive data/information from attack threats.

Each vehicle must first register with the organisation that issues the crucial credentials for communications in order to participate in the network. Once registered, each vehicle develops its own BSM, which contains information about the vehicle, including its speed, real-time location, and direction.

Table I: Existing literature of attack detection in VN

Citation	Year/Author Name	Objective	Future scope
[4]	2021/ Adnan et al.	discussed the vulnerabilities of networks to design a secure and effective architecture for the Software-Defined Vehicular Network (SDVN). To protect communications between V2V and V2I, etc., the Public Key Infrastructure (PKI)-based digital signature system is also utilised.	a suitable method for the attack vector that may necessitate the deployment of reputable resources for remediation, which can consent for investigations and prohibit speedy and secure recovery techniques for restoring the network to a safe operating state, can be made available.
[8]	2019/ Lu et al.	addressed the anonymous security protocols used in VN to safeguard each vehicle's privacy. Next, a few models are discussed along with the key characteristics needed to create effective identity management in VN.	in order for the VNs to provide a reliable communication environment and concurrently safeguard the vehicles' identity and privacy, drivers and passengers will increasingly focus on the reliability of the vast volume of information and demand perfect protection for their privacy.
[9]	2019/ Sheikh et al.	covered the fundamental overview of VN, including their architecture, communication techniques and security services. The discussion of the most recent simulation tools and how well the authentication protocols performed in terms of those tools was followed by a discussion of network applications.	to find malicious activity and defend against all security threats, VN could be equipped with a powerful IDS.
[10]	2020/ Gyawali et al.	to improve accuracy rate and to guarantee the dependability of both vehicles and messages, suggest a reputation- and machine-learning-based misbehaviour detection system (MDS).	Deep Reinforcement Learning will be used to protect VN systems from active and digital threats.
[11]	2018/ Sharafaldin et al.	develop a novel IDS dataset with seven different attack types. The authors took 80 features out of the suggested dataset, choose the best features for evaluation, and then used machine learning methods to calculate the performance measures. By identifying the inaccuracy of other datasets, he also compares the suggested dataset's quality.	desire to expand the number of PCs and carry out more modern attacks.
[12]	2021/ Verma et al.	explored many attack vectors and associated security measures utilising various ML and deep learning (DL) methodologies in VN and also put out a generic architecture for VN.	to provide researchers with future guidance in the paths of VN, the open research issues are gathered.

The connecting cars and RSUs collect these BSMs within their communication range, and they then use them to update the shared database that other RSUs can access. RSUs have a lot of resources, which aids in the detection of attacks. When an attack of any kind is discovered in the network, the RSU broadcasts an alert message to the connected vehicles, while the BS acts as a communication hub for all the RSUs. The normal data is moved to the secure network, which is managed by a trusted organisation.

IV. METHODOLOGY

A. Dataset

The TON-IoT dataset [7]¹ collection includes data from a variety of sources, including network traffic from IoT networks and data from IoT services built by UNSW Canberra's Cyber Range and IoT Labs. In a CSV file, the created data were kept. "Processed datasets" and "Train Test datasets" are the two primary directories for the given datasets. A prepared version of the datasets with their usual characteristics and labels are stored in the "Processed datasets" subdirectory as CSV files. Sample datasets from the "Train Test datasets" folder are used as train-test datasets in a CSV format chosen for testing the effectiveness and correctness of cybersecurity technologies and ML techniques [14]. The proposed dataset

is labelled with a 'label' feature that differentiates between normal and malicious data and a 'type' feature that differentiates between the types of attacks which are ransomware, XSS, backdoors, injections, DoS, DDoS, passwords, scanning and MITM explained below:

- 1) *Backdoor attack*: It allows attackers to quietly get access to the system by defrauding the security protocols and trying to steal sensitive/personal information.
- 2) *Denial of Services attack*: The main purpose of this attack is to shut down the targeted servers/machines, making them inaccessible to their users.
- 3) *Distributed Denial of Services attack*: These attacks occur when a single server is targeted by multiple servers with a DoS attack. In this attack, the target is attacked from many locations at one time.
- 4) *Injection attack*: In these types of attacks, an attacker tries to inject the code or malware into a program on a computer to execute the commands that can change its database.
- 5) *Man in the middle attack*: The major objective of this attack is to steal the person's personal information mainly financial data such as login credentials, account information etc. They spy on personal meetings which contain user's data or secrets.
- 6) *Password attack*: These types of attacks use to take the advantage of your personal information by decoding the

¹<https://research.unsw.edu.au/projects/toniot-datasets>

password. The attacker uses different techniques to access the legitimate user's password.

7) *Ransomware*: It is a type of malware that encrypts the data of a user or any organisation and then demands a ransom payment for the decryption key.

8) *Scanning attack*: In these types of attacks, the attacker scans the devices to collect information such as IP address etc. of these devices before launching advanced attacks to sabotage their security.

9) *Cross-site scripting* : XSS code is illegally injected into a web page or web application. It can compromise the information and the validation process within the IoT devices and a server.

The statistics of the attack category in the proposed TON-IoT dataset is shown in Fig. 2.

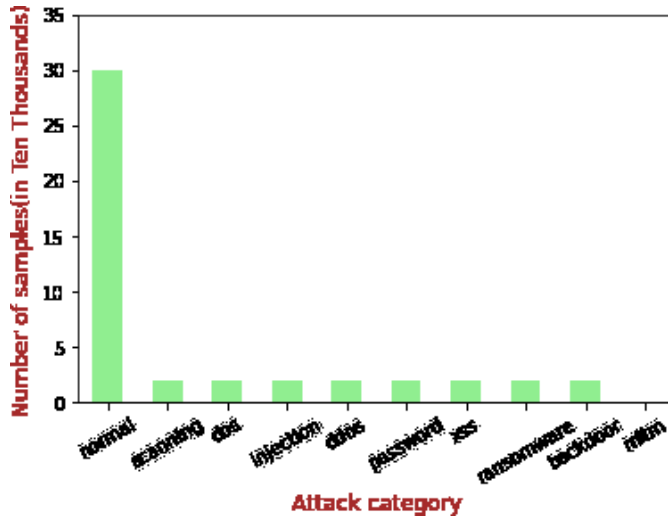


Figure 2: Statistics of attack category

B. Features

The data of the TON-IoT dataset is stored in CSV files containing 46 features. The samples of the datasets in a CSV format are used for the training and testing of the dataset, chosen for measuring the accuracy and effectiveness of the security applications and ML techniques. The dataset includes numerous attacks such as DDoS, password, backdoor, ransomware and MITM, etc. Out of 46 features, some 14 features are extracted to ease the computational process. Table II provides a summary of the TON-IoT dataset's extracted features.

C. Data Preprocessing

For a huge dataset, data preprocessing is an essential step. Before using any ML algorithm [16], the data must be prepared for better accuracy outcomes. This can be done by removing or replacing the missing values if exist and there are some non-numerical features which should be converted into numerical features to conveniently apply the ML methods and the statistical techniques. After the preprocessing of data, it is divided

Table II: Extracted features and their description

Features	Description
ts	Connection timestamp between flow identifiers
src ip	IP address of source
src port	Port number of source
dst ip	IP address of destination
dst port	Port number of destination
proto	Protocols of transport layer in flow of network
src bytes	Source bytes
dst bytes	Destination bytes
src pkts	Source packets
src_ip_byte	Original IP source bytes that is the length of IP header field of source server
dst pkts	Destination packets
dst_ip_byte	Destination IP bytes that is the length of IP header field of destination server
label	label the data as; 0 for Normal data, 1 for malicious data
type	Attack categories

into two parts : Training and Testing. The train-test datasets are structured as CSV files which contain both the training and testing data. First, the dataset is split into two parts: 80% of the data is utilize to train the chosen ML methods, while 20% of the data was saved for the testing dataset.

D. Machine Learning methods

Machine learning is a part of artificial intelligence (AI) that learns patterns from previous data. ML learn to determine various security threats from IoT network that can be applied in detecting network intrusions [17]. On the proposed TON-IoT dataset, the performance indicators are examined using a variety of supervised ML techniques. In particular, three ML methods, namely RF, NB and KNN are used in this paper described below:

1) *Random Forest*: It uses multiple classifiers associated with the data sets to solve complex problems and enhance the performance of the model. RF can be used to analyze an observation based on the outcome of a group of decision trees[18].

2) *Naive Bayes*: It is a probabilistic model based on the Bayes theorem which makes predictions on the ground of probability [7].

3) *K-Nearest Neighbour*: This model can be applied to solve the classification problem. It assumes the identicalness between the new data points and the labelled data points and put the new data points into the group that is more likely the labelled group [14].

These above-mentioned ML models are used to evaluate the performance of the proposed TON-IoT dataset for detecting network intrusions and further making predictions in VN.

V. RESULTS AND DISCUSSION

A. Performance Metrics

The metrics listed below are used to evaluate the effectiveness of several models utilising the TON-IoT dataset to choose the model that best fits this work. These measurements are

calculated using true positive (TP), true negative (TN), false positive (FP) and false negative (FN) explained as:

- **True Positive:** where model predicts the positive instance correctly.
- **True Negative:** where model predicts the negative instance correctly.
- **False Positive:** where model predicts the positive instance incorrectly.
- **False Negative:** where model predicts the negative instance incorrectly.

1) *Accuracy:* It is determined as the number of correct predictions divided by the total number of predictions made. Mathematically, it can be expressed by Eq(1).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

2) *Precision:* It is determined by dividing the total number of correctly predicted instances against the total number of positively predicted instances. Mathematically, it can be expressed by Eq(2).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

3) *Recall:* It is determined by dividing the total correctly predicted instances against the total positive instances. Mathematically, it can be expressed by Eq(3).

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4) *F1-score:* The harmonic mean between the precision and recall is determined by F1-score. Mathematically, it can be expressed by Eq(4).

$$F1_score = \frac{2 * (recall * precision)}{recall + precision} \quad (4)$$

B. Analysis and evaluation

To evaluate the performance of the machine learning models, four parameters accuracy, precision, recall and F1-score are measured using the TON-IoT dataset for the detection of attacks in VN.

Table III: ML models and their accuracy

ML models	Accuracy(%)
RF	94
NB	70
KNN	98.2

Table III shows the accuracy rate of various machine learning models namely RF, NB and KNN. Fig.3 represents the comparison between the ML models, concluding that the KNN model has the highest accuracy rate i.e. 98.2% than the RF and NB models which have 94% and 70% accuracy rates respectively.

Table IV represents the precision(%) of different ML models for intrusion detection using the TON-IoT dataset involving different types of attack. Fig. 4 compares the models about the

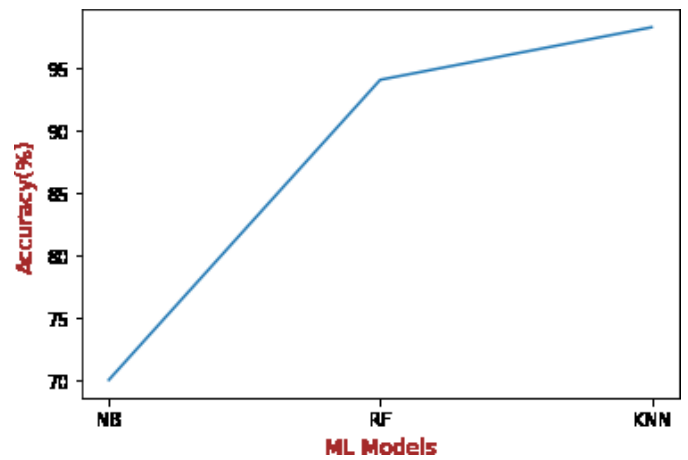


Figure 3: Comparison of ML models with regards to accuracy

Table IV: ML models and their Precision

Attacks	ML Models		
	RF(%)	NB(%)	KNN(%)
Backdoor	94	39	100
DDoS	76	100	100
DoS	100	100	100
Injection	93	100	100
MITM	100	2	98
Normal	96	97	100
Password	68	3	96
Ransomware	99	74	99
Scanning	100	100	100
XSS	91	94	97

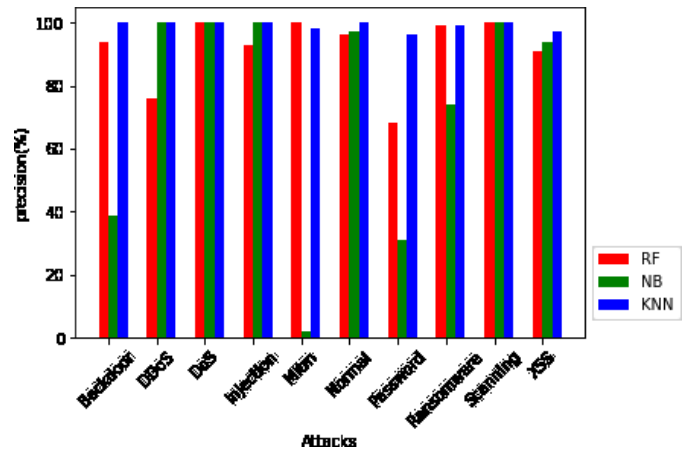


Figure 4: Comparison of ML models with regards to precision

precision where the KNN model gives the highest precision for different types of attacks such as for backdoor, DoS, DDoS, injection and scanning attacks, it gives 100% precision while NB gives the worst precision for the MITM attack i.e. only 2%. RF also give a 100% precision rate for DDoS, MITM and scanning attacks. So, overall KNN gives the best results for precision rate.

Table V represents the recall(%) of different ML models for

Table V: ML models and their Recall

Attacks	ML Models		
	RF(%)	NB(%)	KNN(%)
Backdoor	100	100	100
DDoS	90	100	100
DoS	92	100	99
Injection	98	100	100
MITM	2	97	93
Normal	99	99	56
Password	69	99	100
Ransomware	88	100	85
Scanning	89	100	89
XSS	56	100	100

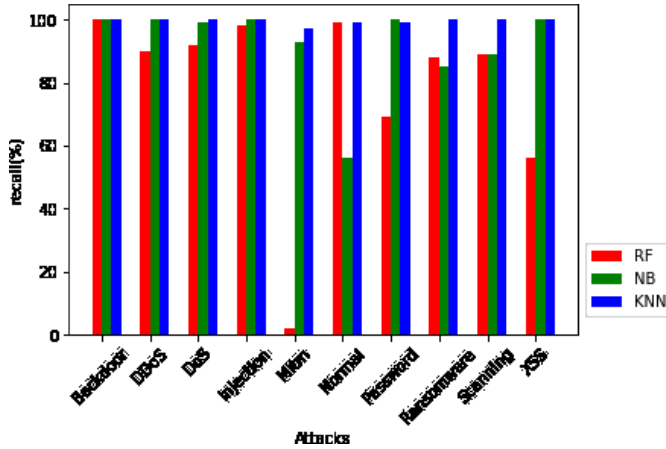


Figure 5: Comparison of ML models with regards to recall

intrusion detection using the TON-IoT dataset which classifies the multiple attacks. RF gives the best recall rate for the backdoor attack i.e. 100% while gives the worst recall rate for the MITM attack i.e. 2%. NB gives a good recall rate for all kinds of attacks as can be seen from Fig. 5 and KNN also gives a better recall rate in comparison to RF model.

Table VI: ML models and their F1-score

Attacks	ML Models		
	RF(%)	NB(%)	KNN(%)
Backdoor	97	56	100
DDoS	83	100	100
DoS	96	100	100
Injection	95	100	100
MITM	3	4	97
Normal	98	71	100
Password	69	48	97
Ransomware	93	79	100
Scanning	94	94	100
XSS	69	97	98

The F1-score(%) of various ML models for intrusion detection using the TON-IoT dataset is represented in Table VI. F1-score combines the performance of two metrics i.e. recall and precision. Fig. 6 compares the models about the F1-score where RF and NB give the worst F1-score for the MITM attack i.e. 3% and 4% respectively. KNN gives the best percentage of F1-score using the ToN-IoT dataset for the attack detection, as seen from Fig. 6.

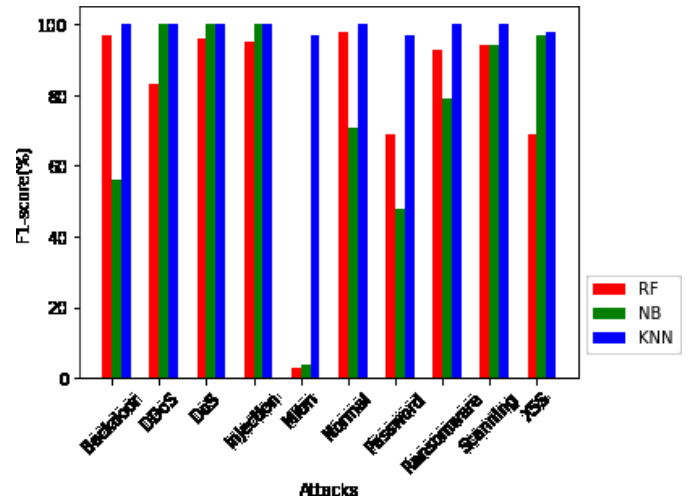


Figure 6: Comparison of ML models with regards to F1-score

VI. CONCLUSION

This paper analyses the TON-IoT dataset for detecting and predicting attacks using the IDS technique in VN. The proposed approach presents a detection framework in the RSUs, which helps in sharing the data with other vehicles and RSUs. The TON-IoT dataset was produced based on a real IoT network environment and displays a wide range of attack categories. A 'label' column is provided to identify both legitimate and malicious data, while a 'type' column is provided to represent the different sorts of attacks. For intrusion detection on a VN, many performance measures like accuracy, recall, F1-score, and precision are measured using ML algorithms. The calculated accuracy rates of RF, NB and KNN are 94%, 70% and 98.2% respectively. The KNN model gives the best accuracy rate using the TON-IoT dataset among all the other models.

REFERENCES

- [1] A. Sharma and A. Jaekel, "Machine learning based misbehaviour detection in vanet using consecutive bsm approach," *IEEE Open Journal of Vehicular Technology*, vol. 3, pp. 1–14, 2022.
- [2] D. Basavaraj and S. Tayeb, "Towards a lightweight intrusion detection framework for in-vehicle networks," *Journal of Sensor and Actuator Networks*, vol. 11, 3 2022.
- [3] A. Talpur and M. Gurusamy, "Machine learning for security in vehicular networks: A comprehensive survey," *IEEE Communications Surveys and Tutorials*, vol. 24, pp. 346–379, 2022.
- [4] M. Adnan, J. Iqbal, A. Waheed, N. U. Amin, M. Zareei, A. Umer, and E. M. Mohamed, "Towards the design of efficient and secure architecture for software-defined vehicular networks," *Sensors*, vol. 21, 6 2021.
- [5] M. Baga, T. Taleb, J. B. Bernabe, and A. Skameta, "A machine learning security framework for iot systems," *IEEE Access*, vol. 8, pp. 114 066–114 077, 2020.
- [6] T. M. Booi, I. Chiscop, E. Meeuwissen, N. Moustafa, and F. T. den Hartog, "Ton-iot: The role of heterogeneity and the need for standardization of features and attack types in iot network intrusion data sets," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 485–496, 2021.
- [7] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. N. Anwar, "Ton-iot telemetry dataset: A new generation dataset of iot and iiot for data-driven intrusion detection systems," *IEEE Access*, vol. 8, pp. 165 130–165 150, 2020.

- [8] Z. Lu, G. Qu, and Z. Liu, "A survey on recent advances in vehicular network security, trust, and privacy," pp. 760–776, 2 2019.
- [9] J. Liang, M. S. Sheikh, and W. Wang, "A survey of security services, attacks, and applications for vehicular ad hoc networks (vanets)," 8 2019.
- [10] S. Gyawali, Y. Qian, and R. Q. Hu, "Machine learning and reputation based misbehavior detection in vehicular communication networks," *IEEE Transactions on Vehicular Technology*, vol. 69, pp. 8871–8885, 8 2020.
- [11] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," vol. 2018-January. SciTePress, 2018, pp. 108–116.
- [12] A. Verma, R. Saha, G. Kumar, and T. H. Kim, "The security perspectives of vehicular networks: A taxonomical analysis of attacks and solutions," *Applied Sciences (Switzerland)*, vol. 11, 5 2021.
- [13] X. Li, Z. Hu, M. Xu, Y. Wang, and J. Ma, "Transfer learning based intrusion detection scheme for internet of vehicles," *Information Sciences*, vol. 547, pp. 119–135, 2 2021.
- [14] N. Moustafa, M. Keshky, E. Debiez, and H. Janicke, "Federated ton iot windows datasets for evaluating ai-based security applications," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 2020, pp. 848–855.
- [15] B. Vivekanadam, "A novel hybrid hnn and firefly algorithm to overcome denial of sleep attack on wireless sensor nodes," *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, vol. 2, no. 04, pp. 223–227, 2020.
- [16] H. Babbar, O. Bouachir, S. Rani, and M. Aloqaily, "Evaluation of deep learning models in its software-defined intrusion detection systems," in *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2022, pp. 1–6.
- [17] M. Hasan, M. M. Islam, M. I. I. Zarif, and M. Hashem, "Attack and anomaly detection in iot sensors in iot sites using machine learning approaches," 2019. [Online]. Available: <https://doi.org/10.1016/j.iot.2019.10>
- [18] M. Sarhan, S. Layeghy, and M. Portmann, "Feature analysis for ml-based iiot intrusion detection," 8 2021. [Online]. Available: <http://arxiv.org/abs/2108.12732>