Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 3 (2022) 499 - 504 ISSN Media Electronic: 2580-0760

Sybil Attack Prediction on Vehicle Network Using Deep Learning

[1]Zulfahmi Helmi, [2]Ramzi Adriman, [3]Teuku Yuliar Arif, [4]Hubbul Walidainy, [5]Maya Fitria*
[1,2,3,4,5]Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala
[5]Telematics Research Center, Universitas Syiah Kuala
[1]zoel.com@gmail.com, [2]ramzi.adriman@unsyiah.ac.id, [3]yuliar@unsyiah.ac.id, [4]hwalidainy@unsyiah.ac.id,
[5]mayafitria@unsyiah.ac.id*

Abstract

Vehicular Ad Hoc Network (VANET) or vehicle network is a technology developed for autonomous vehicles in Intelligent Transportation Systems (ITS). The communication system of VANET is using a wireless network that is potentially being attacked. The Sybil attack is one of the attacks that occur by broadcasting spurious information to the nodes in the network and could cause a crippled network. The Sybil strikes the network by camouflaging themselves as a node and providing false information to nearby nodes. This study is conducted to predict the Sybil attack by analyzing the attack pattern using a deep learning algorithm. The variables exerted in this research are time, location, and traffic density. By implementing a deep learning algorithm enacting the Sybil attack pattern and combining several variables, such as time, position, and traffic density, it reaches 94% of detected Sybil attacks.

Keywords: VANET, Intelligent Transportation System (ITS), Sybil Attack, Deep learning.

1. Introduction

One of the technologies that have been growing rapidly in the Intelligence Transportation System (ITS) is Vehicular Ad Hoc Network (VANET). VANET is a wireless network communication and a subset of Mobile Ad Hoc Network (MANET). This network builds a connection between vehicles and allows communication between them [1]. However, VANET itself is vulnerable to the attack that arises in the networks. One of the attacks that possibly occur in the VANET is the Sybil attack that commonly appears in its routing **[2][3]**. This attack strikes the network by camouflaging itself as an emergency vehicle such as an ambulance, police car, etc., and sharing fake information with the other vehicles [4]. Every node in the VANET communication system can exchange information with another node in the surrounding area **[5]**, e. g. traffic density data that occurs on the road to be traversed by the vehicle. The data is recognized as important information in the system. Furthermore, VANET is also capable of communicating to the network infrastructure, which is the so-called Vehicle to Infrastructure Communication (V2I). V2I enables the vehicle (node) to communicate with traffic lights around and receive information about the duration of a traffic light in that area [6].

In [7], Sudha et. al. predicted the Sybil attack emerges in the VANET by limiting the scope of VANET itself and focusing on the area with high traffic density. The method applied is to make each node a priority in providing information. The Sybil attack takes advantage of this situation to broadcast fake information about the traffic such as sending heavy traffic data where the traffic should be smooth at that time or vice versa. Another study invented by Douceur was conducted to increase the security of the network against the Sybil attack by certifying a trusted network. If there is a new node in the network, the system will identify the node and provide a digital signature to the node by stating that the node is authentic [8]. An algorithm was proposed by Tang and Wang [9] to determine the position or the coordinate of every node in the VANET so that the Sybil attack will be detected conveniently. In addition, Tang proposed the existence of identity (ID) for each node to identify whether the node had previously carried out a Sybil attack or the ID had been forged in a previous Sybil attack.

However, the previous research established in [6]-[9] is only able to detect the Sybil attack by using one variable to obtain the attack pattern. In [6], the attack pattern is detected at a different time, while [7] detected the attack using the wrong coordinate, and the study in [8]-[9]

Accepted: 22-05-2022 | Received in revised: 02-07-2022 | Published: 15-07-2022

utilize the traffic density information to get the Sybil attack pattern. Therefore, this study aims to combine the approach that has been done previously by enacting the attack pattern as a Sybil attack predictive model in VANET. Moreover, this experiment will adopt the neural network method and multilayer from the deep learning algorithm to analyze the input and predict the attack pattern. By using this method, the percentage of early Sybil attack detection in VANET will be increased.

2. Research Method

This research is done by developing the model and simulating the attack in the network using the characteristics of the attack itself. The attack characteristic will be analyzed using a deep learning algorithm to investigate whether the VANET network is infected or not. If the Sybil attack infects the VANET, the system will send a warning before the Sybil could cripple the network.

This experiment is conducted by modeling and simulating the road and vehicles as the nodes in VANET. Trifa et. al. in [8] determined the Sybil attack by restricting the scope of the VANET area to a potential area to be targetted in VANET. One of the areas that are mostly invaded by the Sybil attack is the area with high traffic density. According to Trifa et al., this research will localize the areas that are commonly attacked by the Sybil. The areas simulated are the intersection with five connected roads (Figure 1). The selection of these areas is due to the potential area to be attacked by Sybil and fulfill the aspects that have been inspected by Trifa [8], namely high traffic density, frequent traffic jams, and the fulcrum of vehicle flow within the city.

In Figure 1, it can be seen that the intersection has four roads with two lanes and one-lane road. These four roads head to a roundabout which organizes the flow of vehicles that will change to another lane. The first and the second roads are the access to the office complex, the third road leads to the city center, the fourth road is the access to the culinary, and the fifth road is the access road that leads to the traditional shopping center.

Every road in the intersection has its characteristics such as the time of the traffic jam. In first and second roads will have traffic congestion starting from 07.30 a.m. to 09.00 a.m. At that time the routes will be crowded with vehicles that go to the office or intend to take the kids to the school. At 04.00 p.m., the routes will be back going congested as it is the end time of office hours. The third road which is the city center will be crowded from 04.00 p.m. to 06.00 p.m. The fourth road is a small road whose two lanes will have heavy traffic in the afternoon and the evening. The last road which is the direction to the traditional shopping center will be crowded in the afternoon and on the weekend.

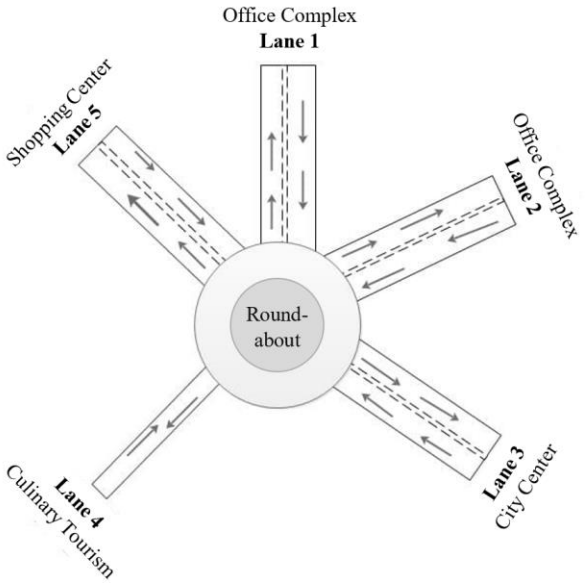

Figure 1. Intersection model of the road [1]

On every road in the intersection, the time and the coordinates of heavy traffic will be noted as a datasheet to be processed with the deep learning algorithm. In addition, traffic support infrastructure mapping such as traffic lights, cyclist paths, etc. is also performed in that area.

The research conducted by Kumar et. al. [7] found that some of the Sybil attacks occur in crowded areas causing severe traffic jams. The congestion causes dense communication data and cripples the VANET network in the area. Another pattern of the Sybil attack that commonly arises is that the attacker broadcasts the fake message to the surrounding nodes. Hence, an authentication ID in each node is required to minimize the attack [8]. Moreover, obstructed data distribution that occurs at one point causes higher throughput and packet loss in the attacked network. According to pattern attack analysis, some variables are considered to influence the Sybil attack, namely the time, the position, and the traffic density.

2.1 The Sybil Attack Variables

Q. Tang et. al. in [9] simulated the Sybil attack by providing false data at different times from its current time. As the result, the VANET network failed to perform data analysis and causes the network system down. Kumar et. al. in [6] proposed a time variable to be an essential variable to detect the Sybil attack. It is done by comparing package receiving time and package delivery time to the node which Sybil is aiming for. The Sybil attack could happen when the package delivery time is too long due to network density. Another simulation by Yao Y. et. al in [10] is conducted by utilizing a device directly in the vehicle to detect the basic method of the Sybil attack. It was found that the Sybil whose more than one node will perform data

DOI: https://doi.org/10.29207/resti.v6i3.4089

transfer in the same period, and continue the transmit to all nodes in the vicinity. Table 1 shows the delay time in the network that is used for sending information from the source node to the targetted node or end-to-end delay. Thus, time is the initial variable used in Sybil attack prediction as the time pattern yield by Sybil is using the same method, and the provider does not detect the time pattern when it is sent. As result, the network could be down.

Tabel 1. End to End Delay Category [8]

Delay Category	Delay value	Index
Excellent	< 150 ms	4
Good	150 – 300 ms	3
Fair	300 – 450 ms	2
Bad	> 450 ms	1

According to Q. Tang et. al. in [9], the ID node is used as attack data. The ID of the node is assigned as an emergency node in the network. In some cases, Sybil provides a fake ID of an emergency vehicle in a certain location. As a result, VANET permits the Sybil node to be the priority node in the network. Thus, Khalil [11] proposed an authentication key to the registered emergency car to the provider in the VANET. The position is decisive for analyzing the ongoing Sybil attack on the network. Information coordinates obtained by Sybil are used as comparison coordinates on the receiving nodes.

The Sybil attack works as if the current situation was the one that happened in a certain area at a certain time. The Sybil attack sends information on the VANET that results in an uncontrolled network in that area. Another pattern of the Sybil attack is by using the identity of the stolen node, such as the identity of an ambulance, fire truck, police car, etc [12]. By having the identity node of emergency vehicles, the Sybil could do anything according to what they aiming for, such as requesting access to use the priority lane in dense traffic. In the case of heavy traffic in a certain area, information will be sent to the Sybil node around and this could cause a low throughput. If this case is left unresolved, it will cause the network to crash and disable the network to forward the information to other nodes. In the end, failure analysis could arise and engender the occlusion and long queue in that lane.

2.2 Simulation Scheme

By using the three variables mentioned, this research combines the attack pattern with the datasheet to be analyzed by applying the deep learning method. There will be time, latitude, longitude, traffic density condition, and traffic light information in the datasheet. In the datasheet, there also will be a table containing the emergency node and its current position.

Table 2 describes the datasheet where the time is changed into an integer. One represents one minute or 60 seconds. This assumption is made to facilitate deep learning in analyzing time data. The position or coordinate is divided into two different columns, namely the longitude and the latitude. This is also to ease the area mapping to the datasheet in deep learning. The condition column in the table represents traffic condition which assumes zero to represent the area is jam-free, and one is when the area is in heavy traffic. Heavy traffic in the system is defined as a long queue 50 meters from the traffic light. The traffic light column is to represent the infrastructure node, e.g. traffic light, in the area. The last column, namely the Is Attack column, points out the given pattern by the time, coordinate, and ongoing congestion in the area. This datasheet will be processed by deep learning as an analysis result for the Sybil attack prediction. Consequently, package data sent by the node and received by another node will be in the form of time, latitude, longitude, traffic condition, traffic light, and the surrounding nodes. The data will then be split and checked in every stage.

Tabel 2. Deep Learning Datasheet

Time	Location		Con-dition	Traffic Light	Is Attack
	Lat	Long			
60	5.556133	95.32162	0	1	0
120	5.556133	95.32162	0	1	0
180	5.556133	95.32162	0	1	0
240	5.556133	95.32162	0	1	0
300	5.556133	95.32162	0	1	0
360	5.556133	95.32162	0	1	0
420	5.556133	95.32162	1	1	0
480	5.556133	95.32162	1	1	1

2.3. Prediction Model of Sybil Attack using Deep Learning

Based on the obtained Sybil pattern, then the prediction model is simulated using deep learning. The deep learning that is implemented in this work utilizes the neural network architecture, which is comprised of interconnected layers, containing an input layer, hidden layers, and an output layer [13].

The simulation of the prediction model is depicted in Figure 2. It can be seen that node 1 is attacked by the Sybil and receives the information to be processed using the deep learning algorithm. The data package sent by the node and received by another node consists of time, latitude, longitude, congestion, traffic light, and surrounding nodes' information. These data later are broken down to be checked in some processes. The simulation step consists of four processes, namely stage 1, stage 2, and stage 3, and prediction results. Figure 3 illustrates how the data is being processed and divided into some layers. In this simulation, received data is in input layers, being processed in three stages of hidden layers, and it is resulting in the prediction model in the output layer.

DOI: https://doi.org/10.29207/resti.v6i3.4089

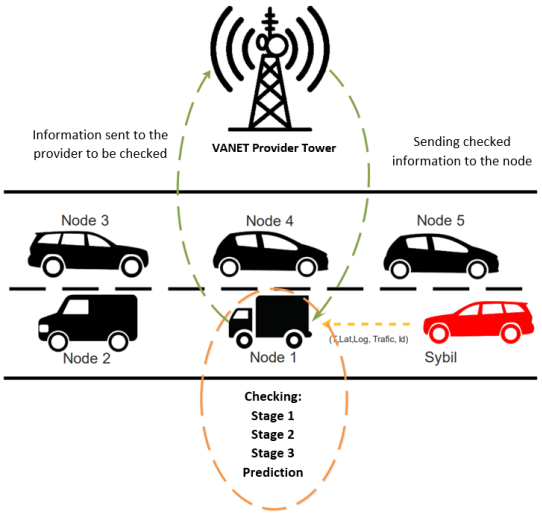

Figure 2. Prediction model of Sybil attack

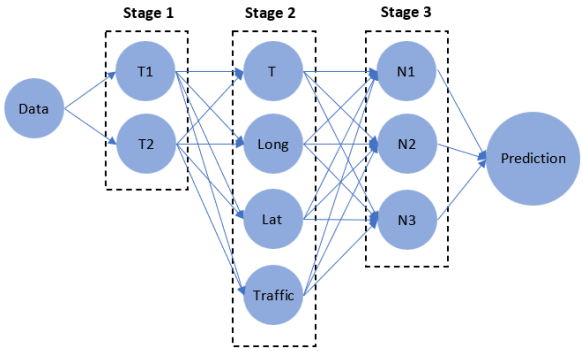

Figure 2. Prediction model of Sybil attack

In stage 1, T_1 and T_2 are used as a part of the simulation. T_1 is obtained from data sent by other nodes, while T_2 is the current time (timestamp) obtained from the recipient nodes. T_1 and T_2 later are compared to yield the time difference. The tolerance time difference is 120 seconds or two minutes [14]. It is considered relevant if the time difference is less than two minutes. Otherwise, the system predicts it as the attack and calculates the percentage of time difference to the current time using the deep learning method.

In the second stage, the time (T), latitude (lat), longitude (long), and traffic information are applied for the matching, coordinate detecting process, and comparing process. Then the comparison values are yielded using the deep learning method. It is assumed that if the deep learning generates a value less than 0.4, then the system will state that the network is not in attack status. On the other hand, if the comparison value is more than 0.4, the prediction status will be marked as attack status.

In this stage, the node will get information about the existence of traffic lights and emergency cars. The recipient nodes will communicate with the VANET provider if there are emergency cars and get the emergency cars' position at the current time. The position data will be compared with the coordinate of the recipient node. If the emergency car position is in its home base or its office, then the status is determined under attack. If the emergency vehicle is around crossroads, then it is stated no attack.

According to the three previous steps applied using the deep learning method, then the average values are calculated. In attack prediction status, a value close to 0 (zero) means that there is no attack (normal) while the value is close to 1 (one) the system will predict that the system is under attack. In this experiment, it is assumed that the value between 0 to 0.39 is the no attack condition, whereas the value between 0.4 to 1 is the condition of the network infected by the Sybil attack.

3. Results and Discussions

The evaluation in this research is conducted within 100 random repetitions. The evaluation data consists of time, coordinate (longitude and latitude), traffic condition, node infrastructure, and emergency node information. Furthermore, some scenarios are designed for evaluation. The first scenario is the evaluation for the attack prediction which is based on the time and traffic conditions. According to 100 tested data, 30 data predicted in normal network status, and 70 others are under Sybil attack.

Table 3. Time and Congestion Datasheet

Time	Traffic Condition	Is Attack
60	0	0
120	0	0
180	0	0
240	0	0
300	0	0
360	0	0
420	1	0
480	1	0
481	1	0
482	1	0

Table 4. Evaluation Data Sample of Time

No	Time	Traffic
1	60	0
2	160	0
3	450	0
4	480	0
5	600	0
6	60	0
7	300	0

After preparing the datasheet of the time and the congestion as in Table 3, then deep learning conducts testing on the datasheet shown in Table 4. As the evaluation finished, the prediction results are provided in Table 5. Table 6 is the simulation results of the Sybil prediction in VANET at the intersection. Subsequently, the evaluation of the Sybil attack scenario is conducted employing the time, coordinate, congestion information, and nearby nodes as served in Table 7.

DOI: https://doi.org/10.29207/resti.v6i3.4089

Table 5. Simulation Results of Time and Congestion

No.	Simulation Results
1	0.0233
2	0.0149
3	0.5339
4	0.7262
5	0.8883
6	0.0233
7	0.0342
8	0
9	0.0188
10	0.0001

Table 6. Comparison of Information Data and Deep Learning Analysis

No	Network Status	Code	Information	Analysis
1	Normal network	N	30	40
2	Network attacked by Sybil	S	70	60
Total Data				

Table 7. Datasheet Using Time, Location, Congestion, and Traffic Light ID

No	Time	Long	Lat	Traffic Condition	Traffic Light
1	60	5.556133	95.32162	0	1
2	160	5.5562773	95.3219835	0	1
3	450	5.5562773	95.3219835	0	1
4	480	5.556173	95.3222348	0	1
5	600	5.55658874	95.32365627	0	1
6	60	5.555805878	95.32205839	0	1
7	300	5.556051482	95.32230515	0	1

Moreover, the evaluation of the prediction process using the deep learning method is also conducted by exploiting the data in Table 8. The evaluation shows all given information by the sender node as in Table 9. The comparison data between information obtained from nodes and the analysis results from deep learning are shown in Table 10 and depicted as a graphic in Figure 3.

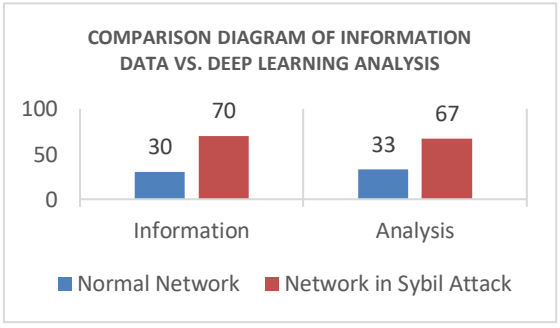

Figure 3. Comparison diagram of information data vs. deep learning analysis using coordinate, traffic, and node

The evaluation is conducted twice by employing 100 data received from the sender nodes. The first evaluation is done by exerting the datasheet consisting of traffic jam information. From that evaluation obtained that 30 of 100 data is in normal network status, while another 70 with network status in Sybil attack as exhibited in Table 5.

Table 8. Datasheet Using Time, Location, Congestion, and Node

No	Time	Long	Lat	Traffic Condition	Traffic Light
1	60	5.556133	95.32162	0	1
2	160	5.5562773	95.3219835	0	1
3	450	5.5562773	95.3219835	0	1
4	480	5.556173	95.3222348	0	1
5	600	5.55658874	95.32365627	0	1
6	60	5.555805878	95.32205839	0	1
7	300	5.556051482	95.32230515	0	1
8	1200	5.555733872	95.32173188	0	1
9	100	5.556083518	95.3215595	0	1
10	1400	5.556281069	95.3217848	0	1

Table 9. Evaluation Results using Time, Location, Congestion, and Node

No	Simulation Results
1	0.0037
2	0.6336
3	0.8782
4	0.9943
5	0.9244
6	0.0037
7	-0.045

Table 10. Results of Information Data vs. Deep Learning Analysis

No	Network Status	Code	Information	Analysis
1	Normal network	N	30	33
2	Network on Sybil Attack	S	70	67
Total Data			**100**	**100**

The evaluation shows that the deep learning method only uses the predictive approach to the predetermined time range. The use of two variables causes less precision in information analysis received by recipient nodes (with an error of 20%). The error is calculated by using formula 1, where SJN is the network difference in normal conditions, while SJS is the network difference in attack status.

$$\% \, error = \frac{SJN+SJS}{All \; data} \quad (1)$$

The evaluation in the second scenario is using time, coordinate, traffic, and node as variables to predict the Sybil attack. In this experiment, the datasheet is prepared are employed as mentioned in the evaluation. The datasheet used is the intersection coordinate (Figure 1), then the time of congestion in the intersection is also taken into account. Fifty coordinates in each intersection area are taken and the time and traffic jam information are set. The evaluation using the deep learning method is performed to get the comparison results between sender nodes and recipient nodes. In this scenario, it is also provided information about the coordinates and time whose traffic jam-free and no infrastructure nodes around. Furthermore, time and congestion evaluations are also be done, but not in the traversed intersection coordinate. The data comparison and deep learning analysis of time, coordinate, traffic, and node are shown in Figure 5. In the Figure, it can be seen that information that is randomly given consists of 30 information in the attack-

DOI: https://doi.org/10.29207/resti.v6i3.4089

free status, and 70 data are in Sybil attack. However, the results obtained by the deep learning method show that 33 data are in normal network status, and 67 others are infected by the Sybil attack.

From the data, it is described that there is an analysis improvement in the results. In the first scenario, the prediction error margin is 20%, while 6% is obtained in the second scenario. This is due to the number of variables compared in the deep learning method.

This experiment indicates that is by combining some variables such as the time, the position, and the traffic condition using the deep learning method generates better Sybil attack prediction. This also simplifies the provider to detect and handle the Sybil attack as soon as possible.

4. Conclusion

Opened communication on VANET has become a security gap for various attacks, one of which is the Sybil attack. The Sybil attack is able to steal the identity of the attacked node and could cripple the network. The network pattern of the Sybil attack can be detected by observing the manipulated time, the incorrect coordinates, and the wrong congestion information. Those patterns are considered to be the variables to determine and analyze the Sybil prediction using deep learning. This study managed to utilize deep learning for detecting the Sybil attack patterns in one VANET area by combining several variables. Implementing layer-by-layer detection, the prediction result achieves 94%. The more variables used in the analysis, the more effective the Sybil predictions. In further research, throughput and delay analysis could deliberately be taken into account to obtain better results in predicting the Sybil attack.

Reference

[1] D. A. Ardiansyah, R. Primananda, and A. Bhawiyuga, "Analisis Kinerja Protokol Routing Ad Hoc on Demand Distance Vector (AODV) pada Jaringan Vehicular Ad Hoc Network (VANET) Berdasarkan Variasi Model Jalan", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 2, pp. 2001-2009.

[2] B. N. Levine, C. Shields, and N. B. Margolin, "A Survey of Solutions to the Sybil Attack", *University of Massachusetts Amherst, MA*, 2006.

[3] N. Balachandran, and S. Sanyal, "A review of techniques to mitigate sybil attacks," *International Journal of Advanced Networking and Applications*, vol. 4, pp. 1–6, 2012.

[4] R. Hadiwiriyanto, P. H. Trisnawan, and K. Amron, "Implementasi Protokol Geographic Source Routing (GSR) Pada Vehicular Ad-Hoc Network (VANET) untuk Komunikasi Kendaraan Dengan Road Side Unit (RSU)", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 12, pp. 7000-7016, 2018.

[5] M. Kabbur, V. A. Kumar, "MAR_Sybil: Cooperative RSU Based Detection and Prevention of Sybil Attacks in Routing Process of VANET", in *Journal Physic: Conference Series,* vol. 1427, no. 1, IOP Publishing, 2020.

[6] R. Anisia, R. Munadi, and R. M. Negara, "Analisis Performansi Routing Protocol OLSR Dan AOMDV Pada Vehicular Ad Hoc Network (VANET)", *Jurnal Nasional Teknik Elektro*, vol. 5, no. 1, pp. 87-97, 2016.

[7] D. Kumari, K. Singh, and M. Manjul, "Performance evaluation of sybil attack in cyber physical system", *Procedia Computer Science*, vol. 167, pp. 1013-1027, 2020. doi: 10.1016/j.procs.2020.03.401.

[8] Z. Trifa, and M. Khemakhem, "Sybil nodes as a mitigation strategy against sybil attack", *Procedia Computer Science*, vol. 32, pp. 1135-1140, 2014. doi: 10.1016/j.procs.2014.05.544.

[9] Q. Tang, and J. Wang, "A secure positioning algorithm against Sybil attack in wireless sensor networks based on number allocating", in *2017 IEEE 17th International Conference on Communication Technology (ICCT)*, pp. 932-936, 2017. doi: 10.1109/ICCT.2017.8359771.

[10] S. Hao, Y. Zhou, and Y. Guo, "A brief survey on semantic segmentation with deep learning", *Neurocomputing*, vol. 406, pp.302-321, 2020. doi: 10.1016/j.neucom.2019.11.118.

[11] M. Khalil and M. A. Azer, "Sybil attack prevention through identity symmetric scheme in vehicular ad-hoc networks," *in 2018 Wireless Days (WD)*, pp. 184-186, 2018. doi: 10.1109/WD.2018.8361717.

[12] S. Moradi, and M. Alavi, "A distributed method based on mobile agent to detect Sybil attacks in wireless sensor networks", in wireless sensor networks. In *2016 Eighth international conference on information and knowledge technology (IKT)*, pp. 276-280, 2016. IEEE. doi: 10.1109/IKT.2016.7777753.

[13] R.M. Cichy, and D. Kaiser, "Deep Neural Networks as Scientific Models", *Trends in Cognitive Sciences,* vo

[14] A. Singh, and P. Sharma, "A novel mechanism for detecting DOS attack in VANET using Enhanced Attacked Packet Detection Algorithm (EAPDA)", in *2015 2nd international conference on recent advances in engineering & computational sciences (RAECS)*, pp. 1-5, 2015. doi: 10.1109/RAECS.2015.7453358.

DOI: https://doi.org/10.29207/resti.v6i3.4089