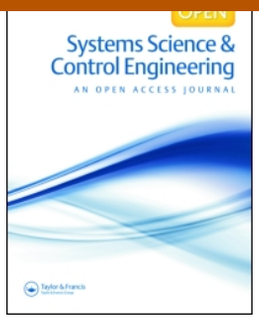

Systems Science & Control Engineering

An Open Access Journal

ISSN: (Print) 2164-2583 (Online) Journal homepage: http://www.tandfonline.com/loi/tssc20

Intelligent intrusion detection in external communication systems for autonomous vehicles

Khattab M. Ali Alheeti & Klaus McDonald-Maier

To cite this article: Khattab M. Ali Alheeti & Klaus McDonald-Maier (2018) Intelligent intrusion detection in external communication systems for autonomous vehicles, Systems Science & Control Engineering, 6:1, 48-56, DOI: 10.1080/21642583.2018.1440260

To link to this article: https://doi.org/10.1080/21642583.2018.1440260

SYSTEMS SCIENCE & CONTROL ENGINEERING: AN OPEN ACCESS JOURNAL, 2018
VOL. 6, NO. 1, 48–56
https://doi.org/10.1080/21642583.2018.1440260

Taylor & Francis
Taylor & Francis Group

Intelligent intrusion detection in external communication systems for autonomous vehicles

Khattab M. Ali Alheeti [a,b] and Klaus McDonald-Maier[a]

[a]Embedded and Intelligent Systems Research Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK; [b]College of Computer Sciences and Information Systems, University of Anbar, Anbar, Iraq

ABSTRACT

Self-driving vehicles are known to be vulnerable to different types of attacks due to the type of communication systems which are utilized in these vehicles. These vehicles are becoming more reliant on external communication through vehicular ad hoc networks. However, these networks contribute new threats to self-driving vehicles which lead to potentially significant problems in autonomous systems. These communication systems potentially open self-driving vehicles to malicious attacks like the common Sybil attacks, black hole, Denial of Service, wormhole attacks and grey hole attacks. In this paper, an intelligent protection mechanism is proposed, which was created to secure external communications for self-driving and semi-autonomous cars. The protection mechanism is based on the Proportional Overlapping Scores method, which allows to decrease the number of features found in the Kyoto benchmark dataset. This hybrid detection system uses Back Propagation neural networks to detect Denial of Service (DoS), a common type of attack in vehicular ad hoc networks. The results from our experiment revealed that the proposed intrusion detection has the ability to identify malicious vehicles in self-driving and even in semi-autonomous vehicles.

ARTICLE HISTORY
Received 1 March 2017
Accepted 9 February 2018

KEYWORDS
Artificial neural networks; intrusion detection system; security; vehicular ad hoc networks; driverless vehicles; semi-autonomous vehicles

1. Introduction

Over the past few years, the traditional internet has expanded to include a ubiquitous network which we refer to as Internet-of-Things (IoTs), which features Machine to Machine (M2M) communication with the ability to provide an efficient connectivity for self-driving vehicles (Zhang, Liang, Lu, & Shen, 2014).

Vehicular ad hoc networks (VANETs) are known as wireless mobile networks which permit self-driving vehicles to easily exchange information such as Cooperative Awareness Messages (CAMs) amongst one another and road side units (RSUs) in their communication area. VANETs are poised to enhance road safety and make service comfortable on busy roads. These networks have the ability to exchange warning messages, notification messages, control data and CAMs between neighbouring vehicles. It is possible for a malicious attack to be launched from any location within the radio area at any time in such wireless networks as they lack a firewalls and gateways. Moreover, a malicious attack does not require physical access to the vehicle, as it would be the case with wired networks (Erritali & El Ouahidi, 2013). Each vehicle can potentially be compromised because they are exposed and the car is allowed to move independently without any physical protection from intrusion (Zhang et al., 2014). Self-driving vehicles rely on participation of other vehicles within the radio coverage area as the external communication uses a decentralized architecture. Potential attackers could thus try to break the supportive protocols between the vehicles and the RSUs. In VANETs, systems security approaches such as encryption/decryption mechanisms and digital signatures can be utilized to decrease the amount of potential attacks, which can be seen as the first layer of defence. Nonetheless, there is a need for a second layer of defence in place for self-driving vehicles to detect and even identify not previously seen attacks (that could not be prevented by the incumbent systems security identified above).

Malicious behaviour in vehicles can be detected initially through the communication of CAMs, via the vehicular ad hoc network between self-driving vehicles and even with their road side units. Every intrusion detection system proposed to protect the routing protocols of VANETs of self-driving vehicles faces to principle problems: (1) providing protection to the transmitted data from one vehicle to the next (i.e. device to device) and (2) protecting and monitoring every CAM which is transmitted between vehicle's road side units. An IDS can be seen as more efficient way to identify an intruder in the external communication of self-driving vehicles (Erritali &

CONTACT Khattab M. Ali Alheeti kmali@essex.ac.uk, kdm@essex.ac.uk, khattabheeti@yahoo.com

El Ouahidi, 2013). The IDS requires the ability to capture and do a thorough examination of every packet which has been received or transferred between vehicles. This process is known as audit data. The security system needs to be able to detect normal and even abnormal behaviours using the audit data which was obtained from VANETs.

An IDS is composed of three phases:

- Data collection phase.
- Analysis data phase.
- Response phase.

In general, two types of detection systems for behaviour security system that are misuse and anomaly detection system. Each of them has advantages and disadvantages which had direct impact on detection performance of any protection system. Misuse detection is a scheme to identifying network and computer attacks. Abnormal behaviour is established in misuse detection to detect any other behaviour is normal. In the same time, abnormal behaviour is defined in anomaly detection to identify from normal behaviour. However, the misuse detection can not detect novel/new attacks, but it is fast and high accuracy with low rate of false alarms. The high rate of false alarms is big disadvantage of anomaly detection in security system.

A hybrid IDS is proposed here for misuse-based detection and anomaly-based detection utilizing Back Propagation artificial neural networks (ANNs) in combination with fuzzy sets, to allow for the prediction of occurrence of attacks on the external communications of self-driving and semi-autonomous vehicles. This IDS should be to detect various types of cyber-attacks like Distributed Denial of Service (DDoS) and network scanning. Additionally, the hybrid security system has the ability to identify and isolate malicious attacks on network. The Kyoto dataset is employed to analyse the performance metrics of the proposed hybrid security system (Song et al., 2011).

The basic structure of communication system in self-driving vehicles is shown in Figure 1. It describes two types of communication in VANETs which are: vehicle to vehicles (V2 V) and vehicle to road side units (V2R). This type of network can provide a communication environment for self-driving and semi-autonomous vehicles that enable to exchange important messages and control data between mobile nodes in that radio coverage area.

The contributions of this research are summarized as follows:

- An intelligent IDS using Back Propagation neural networks to detect abnormal/ malicious behaviours for autonomous and semi-autonomous vehicles.

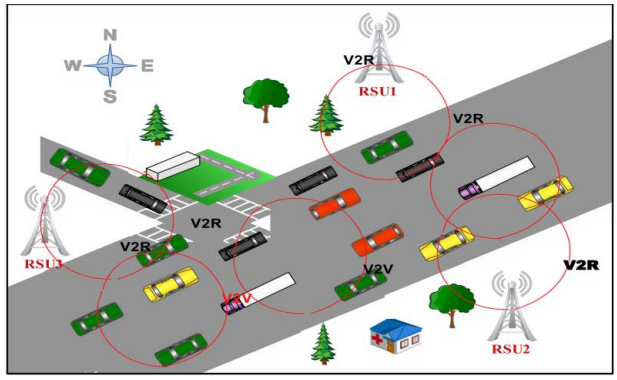

Figure 1. Basic structure of vehicular ad hoc network.

- The proposed intelligent intrusion detection is not dependent on expensive external hardware such as Radar, Lidar, computer vision, or any RSUs.
- The security system is relied on the Proportional Overlapping Scores scheme, which allows to decrease the number of Kyoto benchmark dataset features.
- In this paper, hybrid detection system is proposed to identify Denial of Service that was based on Back Propagation neural networks.

The remainder of the paper is organized as follows: section 2 details the state of the art., section 3 provides the benchmark data set collection. Section 4 presents the proposed protection mechanism, while section 5 provides simulation results. Section 6: discusses the results and section 7 details the conclusion and future work.

2. Literature reviews

It is expected that in the next five years more than 250 million vehicles will be connected with RSUs (Internet of Things & Smart Cities, 2016). Both the self-driving and semi-autonomous vehicles will be very important elements in the IoT domain, with the projection that by 2020 one in every five vehicles will have wireless communication capability (Internet of Things & Smart Cities, 2016).

Petit et al. researched potential threats and cyber-attacks on automation and cooperation automated vehicles (Petit & Steven Shladover, 2015) The result of this research shows that these vehicles can be associated with many different types of attacks. Furthermore, the vehicles need to be created with the awareness of threats in mind and thus this needs to be included during the initial stage development of self-driving vehicles.

In the area of the IoT (Zhang et al., 2014), Zhang et al. offered a survey for defensive schemes and attacks such as Sybil attacks. There were three serious types of Sybil

attacks that were defined by the authors, namely; SA-1, SA-2, and SA-3; these attacks have a direct negative impact on performance and privacy data of smart objects in IoT.

Alheeti et al. proposed intelligent security systems to provide protection to external communication of self-driving vehicles (Alheeti, Gruebler, & McDonald-Maier, 2015). These systems have the capacity to detect both grey hole and rushing attacks utilizing an IDS based on Support Vector Machines (SVM) and Feed-Forward Neural Networks, with SVM being identified as more accurate. Additionally, Tahir et al. provided a security system for the networked sensor systems (Tahir, Tahir, & McDonald-Maier, 2015). In their system, they used the inherent features of a device in order to generate a safe identity for devices in their area.

In (Kang & Kang, 2016), authors utilized deep neural network (DNN) in designing intrusion detection system to secure communication systems of vehicles. The proposed system is heavily based on communication features that are extracted from packets. It has the ability to identify normal and abnormal/ malicious connection. The security system plays important role in protect data that sent/transferred between vehicles in that radio coverage area.

The outlier detection system is proposed to prevent any abnormal behaviour that was has direct and negative impact on performance of any system (Jabez & Muthukumar, 2015). It is heavily based on Neighbourhood Outlier Factor (NOF) in generating anomaly database. A hug dataset is utilized in trained and tested security system to measure its detection accuracy. A novel security system can effectively detect malicious behaviour from others.

Li and et al can find relation between rational cloud resource defender and the potential malicious user (Alrajeh, Khan, & Shams, 2013). This relation plays important role in the cloud as a differential game is investigated. The result of the proposed system can support a theoretical foundation in identifying the abnormal attack. In addition, it can help cloud intrusion detection system in creating optimal decision for systems.

A comparison of various ways which reflects the present case of identification and standardization of open problems of security and privacy of automotive vehicles was made by Petit et al. (Petit, Florian, Michael, & Frank, 2015). This provides a survey of the problems and requirements of vital security in self-driving vehicles. This survey focuses on pseudonym methods, which depend on keys of cryptography mechanisms.

In his previous works, the author researched IDS systems which focused either on detecting misuse (Sato, Yamaki, & Takakura, 2012) or abnormal behaviour (Ahmed, 2009). In this paper, we proposed a first system which address both, the detection of misuse and abnormal behaviour through a hybrid IDS that provides protection and detects attacks in external communication of self-driving and semi-autonomous vehicles.

Table 1. Below shows the type of features.

Feature name	Feature source
Duration, Service, Source bytes, Destination bytes, Count, Same srv rate, SError rate, Srv serror rate, Dst host count, Dst host srv count, Dst host same src port rate, Dst host serror rate, Dst host srv serror rate and Flag	KDD Cup 99′
IDS_detection, Malware_detection, Ashula_detction, Label, Source IP Address, Source Port Number, Destination IP Address, Destination Port Number, Start Time and Duration	Real Network

3. Benchmark data set collection

The data selected for the evaluation of a proposed IDS is an important factor. In the past, researchers used the KDD Cup 99′ data set for analysing the performance of proposed security systems (KDD99-Cup, 1999). Nonetheless, we did not make use of this data in this research work because it suffers from major issues of not providing enough current and recent network topologies and latest attack trends. This data is out of date and based on a system simulation over ten years ago (Song et al., 2011). The Kyoto benchmark is used to carry out the testing and evaluation of performance of the proposed (KDD99-Cup, 1999). It is created for the real traffic data on a network. This important data was obtained over a three-year period. This Kyoto data has two types of data:

(1) Kyoto data set with Internet Protocol (IP) source and IP destination.
(2) Kyoto data set without IP.

We made use of the first option because it uses a label field identifying normal and abnormal connection. It is a Kyoto data set which consists of twenty-four factors which shows normal and abnormal behaviour of a network. We derived the first fourteen features from the previously indicated KDD Cup 99′ data set, while the remaining features were obtained from a real network (Table 1).

Security system researchers recommended the use of this data set in newly proposed IDS′ due to the fact it provides a more practical as well as accurate evaluations and results (Song et al., 2011).

4. The Proposed Intrusion Detection System

The detection system and decreasing the number of features required from the data sets can be seen as the

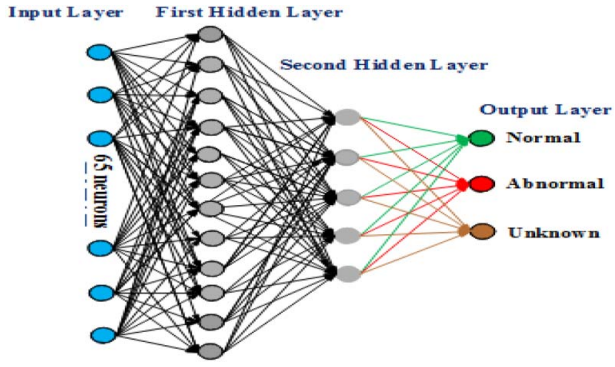

Figure 2. Overall intrusion detection system architecture.

core contributions made in this research work. The number of features utilized has a significant role to play in the effectiveness and the performance efficiency of the proposed security system. Memory requirements, computation time, and accuracy are the key drivers behind the decrease in number of features that are being processed by the system. It has four phases which include;

- The pre-processing phase.
- The features selection phase.
- The fuzzification phase.
- The training and testing phase.

The proposed intelligent IDS uses a Multi-Layer Perception (MLP) which involves four layers: two hidden layers, one output layer and an input layer. The fuzzification data used for the input data of the first layer possesses 65 neurons. The number of neurons and hidden layers depend on the accuracy of the training phase. There are 12 neurons in the first hidden layer while the second one has just five neurons. The proposed security system consists of three values in the output layer which are: unknown, normal, and abnormal.

The following outlines the steps that explain the methodology of the IDS and the how we were able to decrease the number of required features while maintaining the detection accuracy with fuzzification. Figure 2 shows the overall IDS architecture.

4.1. Pre-processing data set phase

We made use of the Kyoto benchmark dataset to analyse the performance of the security system. There is a need for pre-processing stages with this data set, i.e. uniform distribution, encoding, and normalization.

(1) *Encoding stage*: Some features are presented with symbols such as: 'SO', 'REJ', 'RSTO', 'SF'. Before making any changes to the data set, we needed to convert symbol features to numerical values. This is a vital process because the feature vector fed to the input layer of back-propagation (BP) and needs to be numerical.

(2) *Uniform distribution stage:* This is needed with BP to ensure accurate training. We gathered and prepared 60,000 data set records to be used in simulation of the proposed security system. The system dataset is divided into three subsets with each using a different number of normal and abnormal records, randomly obtained from the original dataset. They have the following properties: In a situation where the number of normal pattern is *T.* subset and the original dataset contains *D* samples, then there is a chance of finding a sample of class normal in the first subset *D/T* samples of the final data. Therefore, every subset that belongs to the final data set contains almost the same distribution and ratio of record type of the full data set.

(3) *Normalization of numerical attributes stage:*

According to the equation 1, each numerical is value set between 0.0 and 1.0. ANN training is often more efficient with normalized data; it is used as the preferable predictor.

$$x = \frac{x - \min}{\max - \min} \tag{1}$$

where x is the normalized value with a range between 1 and 0, x is the original value, *max* and *min* are maximum and minimum values of the original variable. These values are utilized in matching the upper and lower limits of the activation function sigmoid which are used in the ANN models. We set some data set aside in the training phase for validation to overcome a common issue in ANN, over-fitting that occurs often during the training phase. This is avoided by stopping the training process when the validation errors increase for a specified number of iterations.

4.2. Extract the impact of features input phase

In the design of the detection system, the feature selection and the ranking process for important features are key (Alheeti, Al-Jobouri, & McDonald-Maier, 2013). Hence, efficient and effective performance of the IDS mostly relies on the number and type of Kyoto features. Alternatively, the removal of less important features enhances the rate of detection, reduces the computation time and memory requirement, therefore improving the overall performance of an IDS. Table 2 shows the effect of the suggested IDS on the time training and consumption of

Table 2. Performance metrics.

Metrics	IDS with all features	IDS with 13 Features
Memory Consumed	72e05b	52e05b
Time	24.31s	21.53s

the memory. Applying 13 features decreases the time needed by 11.4% and the required memory by 27.7%.

The significant features are chosen to raise the detection accuracy and to decrease the volume of the false alarms. A statistical approach is utilized to choose important features which have a high weight value and a critical effect, based on the Proportional Overlapping Scores (POS) technique (Mahmoud et al., 2014).

In order to avoid an outliers effect, the POS is calculated for every feature in Kyoto benchmark. In this situation, the important features selection relies on the measure for the overlap value. The designer can set the dataset size of the selected features (Official site for PropOverlap package, 2016). POS is often considered the most suitable and efficient method among the common types of dataset (Mahmoud et al., 2014). Additionally, it is considered effective even with a dataset which has classification issues like high-dimensional binary and outliers (Official site for PropOverlap package, 2016). POS is used to measure the overlapping rate in the dataset. The recognition features are obtained by calculating the overlap between feature values present in the Kyoto dataset across two classes. The statistical language R is utilized for programming the POS method which pseudocode is presented below:

Proportional Overlapping Scores Algorithm

```
1.  inputs: 'data1.csv'.
2.  output: Sequence of the selected features.
3.  install.packages('propOverlap').
4.  source(' http://bioconductor.org/biocLite.R').
5.  biocLite('Biobase').
5.  library(propOverlap).
6.  ?propOverlap.
7.  getwd().
8.  data < - read.csv('data1.csv',header = T).
9.  str(data).
10. data < - t(data).
11. G < - data[1:23,] # define the features matrix 23.
12. G < - jitter(G). # to avoid the noise in data
13. class < - as.factor(data[24,]) #define class labels.
14. set.seed(1234).
15. selection < - Sel.Features(G, Class, K = 23,Verbose = TRUE) # the
      main function.
16. selection$Features. # extract the number of features
17. selection$Measures. # extract name of features.
```

Each feature is tagged with the overlap value and then the features that bears the lowest weight are removed. The trial-and-error is utilized to determine the optimal number of extracted features based on detection accuracy in the training phase for ANN (Ali Alheeti et al., 2015). Hence, we began with 23 features and we calculated the training accuracy and removed features that had less impact in the detection process after each round of training.

Table 3. Significant features.

Significant feature name	Feature source
Duration, Service, Source bytes, Destination bytes, Count, Dst host count, Dst host srv count, and Flag	KDD Cup 99′
Label, Source IP Address, Source Port Number, Destination IP Address, Destination Port Number and Duration	Real Network

We were able to accomplish 99.18% training accuracy with 13 features which defines the normal and abnormal behaviour in the Kyoto benchmark. Thus the first contribution of this paper is reducing the number of features. The proposed IDS analysed every feature and identified the 13 selected features. These significant features are shown in the Table 3 below.

4.3. The fuzzification phase

Fuzzy sets are seen as a popular solution for a dataset that suffers from classification issues (Ramkumar & Murugeswari, 2014). Choosing the correct features from the Kyoto benchmark dataset has a positive and direct influence on the performance of the proposed IDS (Alheeti, Venus, & Suleiman Al Rababaa, 2009). When the number of features is reduced the classes used to identify normal and malicious behaviours is not clearly separated. The rate of detection in this case, declines and results in an increase of the number of false alarms.

The fuzzification process has the capacity to create a clear border within important features to resolve such classification issues (Ramkumar & Murugeswari, 2014). In Table 4, we have shown the role of the fuzzification in improving results.

$$f(x,a,b,c) = \max(\min(x-a/b-a, c-x/c-b), 0) \quad (2)$$

where, x is the normal value of the dataset before fuzzification while a,b and c values represent the fuzzy domain values. The proposed security system is ultimately more efficient with fuzzification data (Ali Alheeti, Gruebler, & McDonald-Maier, 2015). Thus, it has the ability to overcome the confusion or ambiguity through redistributing each feature value with new five values. Equation 2 allows each value from the selected features to take five values from the fuzzy domain with interval range is [0, 1].

Table 4 shows comparative analysis results for various configurations of the IDS. We analyse the IDS against a full set of features. We also analysed it against a decreased set of features to assess the performance of the system.

Table 4. Performance metrics.

Metrics	IDS with all Features	IDS with 13 Features	IDS with 13 Fuzzification Features
Misuse Detection Normal	97.5%	99.79%	99.23%
Misuse Detection Abnormal	99.2%	64.34%	99.05%
Anomaly Detection Normal	92.04%	60.35%	99.04%
Anomaly Detection Abnormal	99.85%	98.45%	99.06%
Unknown Rate	28.5	0%	0.03
Average FP Alarm	2.27%	23.61%	1.82%
Average FN Alarm	1.01%	7.38%	0.4%
Average Error Rate	1.9%	19.32%	0.88%
Training Parameter Epochs	115	75	27

Table 4 also highlights the vital role of the fuzzification dataset in improving the rate of detection, reducing the amount of false alarms and error rate. Additionally, fuzzification features have a positive reflect on the training phase for ANN by reducing the number of epochs.

4.4. The training and testing phase

The IDS, uses a well-known supervised learning of neural network architecture known as Multi-Layer-Perceptron (MLP) with back back-propagation gradient-descent in the designed security system (Alheeti et al., 2009). In order to form a feed-forward multi-layer in MLP, the collection of non-linear neurons is connected to one another (Al-Naqshabandi, 2007). This technique is known to be very useful for prediction and classification issues. Figure 3 illustrates the graphical representation of the initial MLP employed in this research. Cross-validation is used to determine the 'optimal' number of hidden layers and neurons which were relied on the experimental design of the IDS. Particularly, the training of the MLP began from a small number of neurons, and with only one hidden layer, which measures the error ratio of the trained BP on holdout samples, slowly increasing the number of neurons at the hidden layer in which the performance of the trained phase on holdout samples has begun to go down due to the problem of overtraining. Thus, we obtain the best number of neurons for the hidden layer of the ANN.

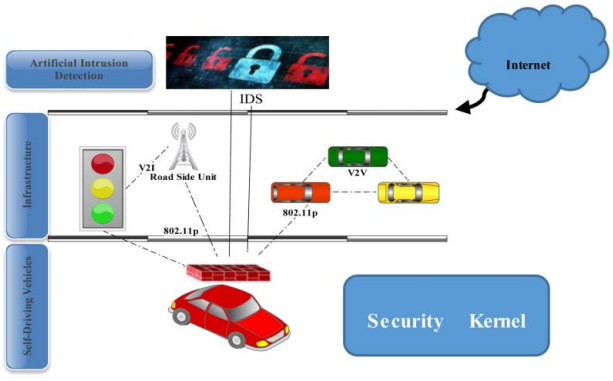

Figure 3. Graphical representation of our MLP neural network model.

We employed a 100-fold cross-validation to reduce the bias related with the process of randomly separating the Kyoto data set into testing phase and training phase. The network training is ended when the Least-square-error E between the desired d_i and actual outputy_i is less than E_{max} or when the number of sweeps equal 500. We define $E_{max=}1 * 10^{-7}$.:

$$E = \frac{1}{2_P} \sum_{P=1}^{P} \sum_{i=1}^{m} (y_i - d_i)^2 \tag{3}$$

where p is the total number of training patterns, and:

$$d_i = \begin{bmatrix} 1 & \text{If the training pattern} & \in i^{th}\text{cluster} \\ -1 & & \text{otherwise} \end{bmatrix}$$

For all experiments, the learning rate α was fixed to $1 * 10^{-7}$. A trial-and-error approach is employed to iteratively get the optimal training accuracy rate. Table 5 illustrates some of the configuration parameters employed in the ANN.

In the following phase, the detection phase, we tested the security system with important features that we chose from the Kyoto benchmark data set. We analysed their behaviours and used the IDS to generate four types of alarms: true negative, true positive, false negative and false positive (Alheeti et al., 2015). These rates of detection accuracy and alarms are used to calculate the performance of IDS. We have three outputs in the detection phase which includes: unknown, normal, and abnormal. We created the simulation on a system making use of an Intel core i3 processor operating at 2.53 GHz and 4GB RAM memory. In addition, the proposed security system is shown in Figure 4 that has been configured on each vehicle to identify abnormal activates on the external compunction system in self-driving vehicles.

Table 5. Artificial neural network parameters.

Parameter	Value
Train Parameter epochs	81
Train Parameter ln.	$1*10^{-7}$
Train Parameter goal	0
Train Parameter Minimum Gradient	$1*10^{-11}$

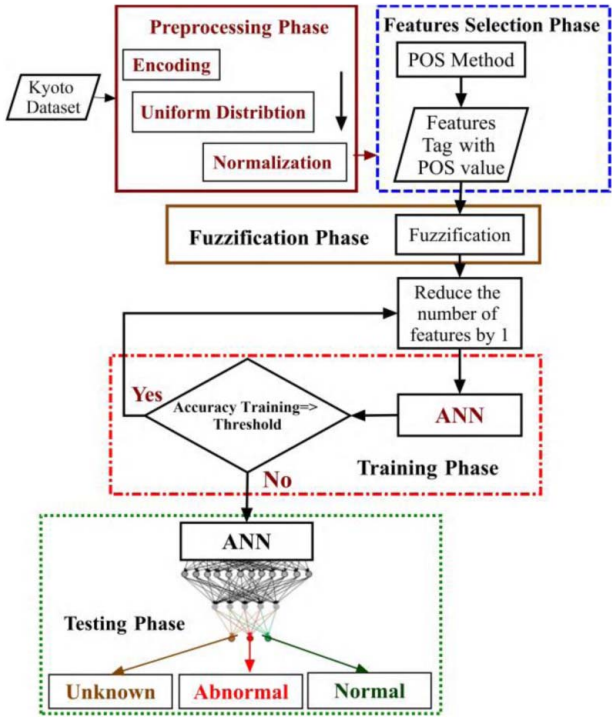

Figure 4. System architecture.

We categorized the test dataset into three categories in the testing phase which includes unknown, the normal and the abnormal. We employ the misuse and anomaly detection so that our ANN can efficiently learn the small and malicious behaviours via the iterative operation (Using Artificial Intelligence to create a low cost self-driving car, 2016).

In this paper, a hybrid intelligent detection system is designed to secure external communication systems of autonomous vehicles. It was different from the previous security systems by which it has ability to detect a novel intruder. Moreover, the proposed security system can reduce the number of features that have been extracted from the Kyoto dataset. Therefore, the number of those features has been distributed by utilizing fuzzy sets which has an important role in improving accuracy of detection system and reducing the rate of alarms. The intelligent hybrid security system that proposed in this research has the ability to detect abnormal behaviours for self-driving vehicles. In addition, it has the ability to protect the external communication system of autonomous vehicles.

5. Experimental evaluation and results

The proposed protection mechanism employed a dataset of 60,000 records to define the normal, abnormal and unknown behaviour in networks. The data set employed in our analysis was obtained from the Kyoto University, as shown above (Koyoto dataset, 2016, April). The accuracy of the training algorithm is = 99.18%. The accuracy of the detection is calculated based on Equation 4 below (Alheeti et al., 2009):

$$\text{Accuracy} = \frac{\text{Number of correctly classified patterns}}{\text{Total number of patterns}} \quad (4)$$

$$TP_{\text{Rate(sensitivty)}} = \frac{TP}{TP + FN} \quad (5)$$

$$TN_{\text{Rate(specificity)}} = \frac{TN}{TN + FP} \quad (6)$$

$$FN_{\text{Rate}} = (1 - \text{sensitivity}) = \frac{FN}{FN + TP} \quad (7)$$

$$FP_{\text{Rate}} = (1 - \text{specificity}) = \frac{FP}{FP + TN} \quad (8)$$

The data set was divided into three subsets: the test set (25%), the validation set (25%) and the training set (50%).

5.1. Training and testing IDS with misuse detection

We employed the misuse scheme in creating the IDS. It has two features, which are low false alarms and high detection rate. The number of records and the accuracy of the detection rate that were employed in our IDS is illustrated in Tables 6 and 7 demonstrates the rates of alarms of the proposed IDS.

The alarm rate and error rate that are generated from the security system are shown in Table 7. Equations (5–8) are utilized in calculating the alarm rate.

5.2. Training and testing IDS with anomaly detection

The training and testing of the proposed security system are performance with the Kyoto data set (KDD99-Cup, 1999). We discovered that the proposed IDS performance is directly related to anomaly detection algorithm. In a

Table 6. Classification rate.

Class	Original No.	Neural No.	Accuracy
Normal	6895	6829	99.04%
Abnormal	3105	3076	99.06%
Unknown	0	4	0.04%

Table 7. Alarm and error rate.

Alarm Rates	
True positive	99.59%
True negative	97.99%
False negative	0.40%
False positive	2.00%
Error Rate	0.95%

Table 8. Classification rate.

Class	Original No.	Neural No.	Accuracy
Normal	6640	6887	99.23%
Abnormal	3060	3031	99.05%
Unknown	0	3	0.03%

Table 9. Alarm and error rate.

Alarm Rates	
True positive	99.60%
True negative	98.34%
False negative	0.40%
False positive	1.65%
Error Rate	0.82%

case were the anomalies are correctly detected from the data set, it offers a high detection rate and also less false alarms. The detection of the anomaly provides the capability to identify novel attacks. Our experiment shows that the IDS performance is largely impacted by the type of the training data.

The system measured the classification rate and generated four types of alarms for the suggested IDS as illustrated in Table 8.

Table 9 demonstrates the rate of four alarms and error rate of detection system of communication of self-driving vehicles.

Here we present a hybrid intelligent detection system that was different from the previous intrusion detection systems by which it has ability to identify a novel attack. However, the proposed system can reduce the number of features that have been extracted from the Kyoto benchmark dataset. Moreover, the number of those Kyoto features has been distributed by utilizing fuzzy sets which has an important role in improving accuracy of detection system and reducing the rate of alarms. The intelligent hybrid security system that proposed in this paper has the ability to detect malicious behaviours for self-driving vehicles. In addition, it has the ability to secure the external communication system of autonomous vehicles.

6. Discussion

Conventional security systems do not have sufficient capacity to provide adequate security to the external communication of self-driving and semi-self-driving vehicles. There is a serious need to improve security in order to ensure functionality and quality of service in these types of networks.

The rate of alarm in our paper lays between 97.99% and 99.60%. This facilitates an efficient rate of detection with an average rate error of 0.88%, while the previous best accomplished average error rate was 8.68% (Ali Alheeti, Gruebler, & McDonald-Maier, 2015). In (Ali Alheeti et al., 2015), the average rate of false alarms is at 4.86%, while we accomplish 1.64% presented here. Thus the hybrid IDS performance is efficient in identifying DoS of communicating self-driving vehicles. With this result, we can safely decrease the number of features and fuzzification data in enhancing the performance of the security system.

Furthermore, we can overcome overlapping problems with the proposed IDS by using the POS approach in order to choose the significant features and apply fuzzy set 'fuzzification' to the features that we extracted. Intelligent IDS can provide protection to the VANETs of self-driving vehicles by identifying and blocking novel behaviours which target vehicles by affecting the communication between self-driving and semi self-driving vehicles. We can observe the vital role of the IDS from our experiment in enhancing the external security of communication vehicles under various conditions. The IDS directly affects the performance of the network by enhancing the rate of detection, as well as reducing the number of false alarms and the error rate.

In this paper, the proposed system is evaluated with new dataset to test the detection performance with an other dataset. Assessing the detection performance of the proposed IDS with a new dataset is important to check the efficiency of the protection system. The new dataset is extracted from trace file of ns-2. The normal and abnormal behaviours are generated to test the proposed security system.

7. Conclusion and future work

In this work, we have been able to design a hybrid IDS which combines MLPs with POS and fuzzy sets to identify the behaviour of connected communicating vehicles. The proposed intrusion detection system employs an intelligent misuse and anomaly detection method in identifying the malicious behaviour in VANETs of self-driving vehicles which safely adapts to heterogeneous communicating environment.

The security system has the capacity to identify the malicious vehicle which is a source of an attack by employing the hybrid IDS. It can address the issues of malicious intrusion in self-driving vehicles. This scheme was created for the testing phase and the training phase of two system scenarios: normal and abnormal behaviour which is based on the Kyoto data set. The IDS is employed to evaluate the behaviour of each vehicle to identify if it is a vehicle causing DoS or just a normal vehicle. In a situation where the car is trying to prevent the access to the resources of the network and render them unavailable at any time, this is shown as a DoS vehicle. The hybrid proposed BP-IDS has the capacity to differentiate between existing, novel or new attacks.

The experiment shows IDS performance in identifying and isolating DoS. It is good for protecting the external communication of self-driving vehicles. Additionally, it possesses the capability to identify external and internal attacks launched on the network at any time.

With our proposed IDS, malicious self- driving vehicles can be isolated with high rate of detection and can guarantee low false alarm. The process of reducing the features by POS technology has an important role to play in improving the rate of detection of the proposed IDS. Additionally, the fuzzification data help reduce the amount of false alarms and error rate when compared with our previous publication. In future work, there is a need to use the IDS on VANETs based utilizing a virtual communications layer, and we anticipate a further improved intrusion behaviour.

Acknowledgements

This work has been supported by the UK Engineering and Physical Sciences Research Council EPSRC [EP/K004638/1, EP/R02572X/1 and EP/P017487/1].

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Khattab M. Ali Alheeti http://orcid.org/0000-0002-6393-7410

References

Ahmed, B. (2009). *Link analysis approach to improve detection of fragmentation attacks in misuse IDS*. 2009 first international conference on communications and networking, (pp. 1–8).

Alheeti, K. M., Al-Jobouri, L., & McDonald-Maier, K. (2013). *Increasing the rate of intrusion detection based on a hybrid technique*. 5th computer science and electronic engineering conference, (pp. 179–182), 2013.

Alheeti, K. M., Gruebler, A., & McDonald-Maier, K. (2015). *On the detection of grey hole and rushing attacks in self-driving vehicular networks*. Computer science and electronic engineering conference (CEEC), 7th 2015 Sep 24, (pp. 231–236).

Alheeti, K. M., Venus, W., & Suleiman Al Rababaa, M. (2009). *The affect of fuzzification on neural networks intrusion detection system*. Industrial electronics and applications, ICIEA 2009. 4th IEEE conference on 2009 May 25 (pp. 1236–1241).

Ali Alheeti, K., Gruebler, A., & McDonald-Maier, K. D. (2015). *An intrusion detection system against black hole attacks on the communication network of self-driving cars*. 2015 sixth international conference on emerging security technologies (EST), 2015 September 3 (pp. 86–91).

Ali Alheeti, K., Gruebler, A., & McDonald-Maier, K. D. (2015). *An intrusion detection system against malicious attacks on the communication network of driverless cars*. Ieee consumer communications and networking conference (CCNC), 12th annual 2015 (pp. 916–921).

Al-Naqshabandi, S. M. (2007). *Simulation system for computer network intrusion detection*. A thesis submitted in partial fulfillment of there quirements for the degree of doctor of philosophy in computer science (pp. 61–66), Al-Nahrain University, Baghdad, Iraq.

Alrajeh, N. A., Khan, S., & Shams, B. (2013, May). Intrusion detection systems in wireless sensor networks: A review. *International Journal of Distributed Sensor Networks*, *9*(5), 167575.

Erritali, M., & El Ouahidi, B. (2013). A review and classification of various VANET intrusion detection systems. *IEEE Security Days (JNS3)*, *2013*, 1–6.

Internet of Things & Smart Cities. (2016, January). Retrieved from http://Downloads/151102_insights_capitalising_on_internet_of_things

Jabez, J., & Muthukumar, B. (2015, January). Intrusion detection system (IDS): anomaly detection using outlier detection approach. *Procedia Computer Science*, *48*(48), 338–46.

Kang, M. J., & Kang, J. W. (2016, June). Intrusion detection system using deep neural network for in-vehicle network security. *PloS one*, *11*(6), e0155781.

KDD99-Cup. (1999). The third international knowledge discovery and data mining tools competition dataset KDD99-Cup. Retrieved from http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

Koyoto dataset. (2016, April). Retrieved from http://www.takakura.com/kyoto_data/BenchmarkData-Description-v5

Mahmoud, O., Harrison, A., Perperoglou, A., Gul, A., Khan, Z., Metodiev, M., & Lausen, B. (2014). A feature selection method for classification within functional genomics experiments based on the proportional overlapping score. *BMC Bioinformatics*, *15.1:274*, 1–20.

Official site for PropOverlap package. (2016, April). Retrieved from http://cran.r-project.org/web/packages/propOverlap/index.html

Petit, J., Florian, S., Michael, F., & Frank, K. (2015). Pseudonym schemes in vehicular networks: A survey. *Communications Surveys & Tutorials*, *17*(1), 228–255.

Petit, J., & Steven Shladover, E. (2015). Potential cyberattacks on automated vehicles. *Intelligent Transportation Systems, IEEE Transactions on*, *16*(2), 546–556.

Ramkumar, J., & Murugeswari, R. (2014). *Fuzzy logic approach for detecting black hole attack in hybrid wireless mesh network*. 2014 ieee international conf. On innovations in engineering and technology (ICIET'14), Vol. 2347–6710, (pp. 877–882).

Sato, M., Yamaki, H., & Takakura, H. (2012). *Unknown attacks detection using feature extraction from anomaly-based ids alerts*. Applications and the internet (SAINT), 2012 IEEE/IPSJ 12th international symposium on, (pp. 273–277), .

Song, J., Hiroki, T., Yasuo, O., Masashi, E., Daisuke, I., & Koji, N. (2011). *Statistical analysis of honeypot data and building of kyoto 2006+ dataset for NIDS evaluation*. Proceedings of the first workshop on building analysis datasets and gathering experience returns for security, (pp. 29–36), ACM.

Tahir, H., Tahir, R., & McDonald-Maier, K. (2015). *Securing MEMS based sensor nodes in the internet of things*. 2015 ieee sixth international conference on emerging security technologies (EST), (pp. 44–49).

Using Artificial Intelligence to create a low cost self-driving car. (2016, April). Pdf.

Zhang, K., Liang, X., Lu, R., & Shen, X. (2014). Sybil attacks and their defenses in the internet of things. *IEEE Internet of Things Journal*, *1*, 372–383.