Title: A credibility score algorithm for malicious data detection in urban vehicular networks

Author: Bartłomiej Płaczek, Marcin Bernas, Marcin Cholewa

Citation style: Płaczek Bartłomiej, Bernas Marcin, Cholewa Marcin. (2020). A credibility score algorithm for malicious data detection in urban vehicular networks. "Information" Vol. 11, iss. 11 (2020), art. no. 496, doi 10.3390/info11110496

information

MDPI

Article

A Credibility Score Algorithm for Malicious Data Detection in Urban Vehicular Networks

Bartłomiej Płaczek [1,*], Marcin Bernas [2,*] and Marcin Cholewa [1]

[1] Institute of Computer Science, University of Silesia, Będzińska 39, 41-200 Sosnowiec, Poland; marcin.cholewa@us.edu.pl

[2] Department of Computer Science and Automatics, University of Bielsko-Biała, ul. Willowa 2, 43-309 Bielsko-Biała, Poland

* Correspondence: placzek.bartlomiej@gmail.com (B.P.); marcin.bernas@gmail.com (M.B.); Tel.: +48-32-368-9760 (B.P.)

Received: 29 September 2020; Accepted: 19 October 2020; Published: 23 October 2020

Abstract: This paper introduces a method to detect malicious data in urban vehicular networks, where vehicles report their locations to road-side units controlling traffic signals at intersections. The malicious data can be injected by a selfish vehicle approaching a signalized intersection to get the green light immediately. Another source of malicious data are vehicles with malfunctioning sensors. Detection of the malicious data is conducted using a traffic model based on cellular automata, which determines intervals representing possible positions of vehicles. A credibility score algorithm is introduced to decide if positions reported by particular vehicles are reliable and should be taken into account for controlling traffic signals. Extensive simulation experiments were conducted to verify effectiveness of the proposed approach in realistic scenarios. The experimental results show that the proposed method detects the malicious data with higher accuracy than compared state-of-the-art methods. The improved accuracy of detecting malicious data has enabled mitigation of their negative impact on the performance of traffic signal control.

Keywords: vehicular network; vehicular ad-hoc network (VANET); malicious data; traffic signals; smart city; intelligent transportation systems

1. Introduction

A specific feature of road networks in urban areas is the necessity of controlling traffic lights at signalized intersections. The recent development in wireless vehicular network technology has enabled the introduction of new systems that are using information delivered from particular vehicles to control the traffic lights. In such systems, the vehicles send their current position (speed, etc.) to a road-side control node. These data are used in real time to optimize traffic signals, i.e., to adapt the duration of signal phases and minimize delays of the vehicles that have to stop at intersection. By taking into account the detailed data from vehicles, the control node can better adapt traffic signals to actual needs of traffic participants. Thus, a higher performance of the traffic signal control can be achieved, which means reduced vehicle delays, lower fuel consumption, emission, and congestion as well as improved safety [1,2].

The effective operation of traffic signal control, based on data from vehicular network, is impossible in case of presence of malicious data. For instance, the malicious data can be injected in the vehicular network by a selfish vehicle approaching an intersection. Such vehicle can perform Sybil attack, i.e., report

multiple false vehicles in a given traffic lane, to get the green light immediately. Another source of the malicious data can be vehicles with malfunctioning sensors, e.g., a moving vehicle, which continuously reports the same (false) position. Results of previous study have shown that a low fraction of malicious data can significantly impact the effectiveness of traffic signal control [3]. Therefore, it is necessary to detect and filter out the malicious data.

This paper introduces a method to recognize the malicious data in the urban vehicular networks, that are composed of vehicle nodes and road-side units. The considered scenario assumes that the vehicle nodes periodically transmits their position to the road-side units controlling traffic signals at intersections. Detection of the malicious data is conducted with use of a microscopic traffic model based on cellular automata. This traffic model allows to determine intervals representing possible positions of vehicles approaching the signalized intersections. A credibility score algorithm is introduced to decide if positions reported by particular vehicles are reliable and should be taken into account for controlling traffic signals. A realistic simulation environment was used to verify if the proposed approach accurately detects the malicious data and ensures effective operation of the traffic signal control system.

The rest of this paper is structured as follows. Related works and authors' contributions are discussed in Section 2. Section 3 introduces the proposed method and the credibility score algorithm. Experiments and their results are described in Section 4. Finally, conclusions are given in Section 5.

2. Related Works and Contribution

The problem of malicious data detection in vehicular networks has gained a considerable interest from the research community. Various methods have been proposed so far to recognize and eliminate the malicious data [4–6]. The state-of-the-art methods can be grouped into four categories that include approaches based on behaviour, trust, consistency, and plausibility [7]. This section discusses the most important features of the above mentioned approaches and reviews the most recent related works. A detailed survey of the state-of-the-art methods can be found in [7,8].

The behaviour-based approaches have been proposed to detect malicious information regarding traffic events, such as accidents or congestion. For instance, false accident information can be detected by comparing actual trajectory of the reporting vehicle with an excepted trajectory, that is determined assuming typical driver behaviour in case of accident [9]. If these two trajectories are significantly different, then the event information is recognized as malicious. Another approach is to compare the event information with behaviour of other drivers in vicinity of the reported event [10].

The trust-based methods require data sharing between nodes in the vehicular network to determine a level of confidence of one vehicle node based on information reported by other vehicle nodes in a neighbourhood [11]. An example of the trust management approach is the algorithm presented in [12]. According to this algorithm, each vehicle receives a beacon message from a neighbour, which includes speed and density information. The trust level is determined by taking into account difference between the received information and a threshold value. Similarly, position information transmitted by one vehicle can be verified by other nodes in the vicinity, i.e., by neighbouring vehicles or by roadside units. This approach was implemented in [13] by using the sensors that are installed in vehicles as a part of advanced driver-assistance systems. In case of low traffic density, the trust-based methods suffer from a lack of enough information about the neighbouring nodes [14]. Moreover, the exchange of messages between nodes is a time-consuming operation.

The basic idea behind the consistency approach to malicious data detection is to verify if messages received from vehicles, in successive time steps, are consistent with each other. For instance, when one vehicle is taken into account, then the distance between positions of the vehicle, reported at time steps t and $t+1$, has to be consistent with the speed indicated in message sent at time step

$t+1$. The consistency algorithms also try to resolve conflicts between received inconsistent data, where possible [15]. Other methods check the consistency for data collected from multiple vehicles. In this case previous average speed of neighbouring vehicles must be consistent with the new reported speed. In [16] the Kalman filter is used to track the mobility information received from neighbouring vehicles and error of the Kalman filter is utilized to evaluate data consistency for each vehicle in the neighbourhood. Another approach is to verify if traffic data transmitted by a group of vehicles, from a specified region at a given time, is consistent with the fundamental properties described by traffic flow theory [17]. This method estimates headway and speed distributions based on fundamental dependency between traffic flow and density. Malicious data in vehicular networks are detected, when a high deviation is encountered from expected average headway and speed of vehicles.

The plausibility-based methods rely on physical rules and models (e.g., kinematic models, the relation between time, distance, and speed) to explain the validity of the data [18]. Basic plausibility verification methods detect the malicious data by using some predefined rules, e.g., two vehicle nodes cannot simultaneously occupy the same location by two nodes at the same time. The speed of vehicle nodes should not exceed a maximum speed limit. More sophisticated methods combine knowledge of road geometry and vehicular dynamics. In [19] the plausibility of position, reported by vehicle node, is verified using simulations based on physics of vehicular motion of the transmitting vehicle and physical characteristics of the road such as curvature, length and traction. Disadvantage of such methods is the high computational complexity.

Other interesting approaches to malicious data detection include the application of unmanned aerial vehicles [20] and blockchain algorithm [21]. These methods require special equipment and lead to increased transmission overhead.

The related works are focused on safety and traffic monitoring applications of the vehicular networks. In contrast, this work is devoted to malicious data detection in urban vehicular networks, that include road-side units controlling traffic signals at intersections. The main objective here is to eliminate the malicious data, which can reduce the effectiveness of traffic signal control. According to the authors' knowledge, the above-mentioned problem has not been investigated so far by other researchers.

In our previous paper [3] we have introduced a method of malicious data detection in vehicular networks intended for traffic signal control systems. The previous method combines state-of-the-art approaches (a model of expected driver behaviour with position verification) and assumes that vehicles are equipped with sensors capable to detect neighbouring cars. The experiments, in that study, were conducted using a simple model of road network with single-lane unidirectional roads.

Our main contribution in this paper is a novel credibility score algorithm for detection of the malicious data in urban vehicular networks. The algorithm evaluates credibility of the collected data by using a microscopic traffic model. The traffic model describes mobility of individual vehicles in urban road network. It enables estimation of intervals that reflect expected uncertain positions of particular vehicles in signalized arterial roads. Details of the proposed algorithm are discussed in Section 3. It should be noted that in traffic signal control systems the malicious data have to be filtered in real time. Therefore, to eliminate transmission delay, the proposed approach avoids data exchange between vehicles. The introduced solution is cost effective and easy to implement as any additional sensors do not have to be installed in vehicles. The contribution of this work also includes experimental evaluation of the proposed method in realistic simulation environment for complex road network and various traffic control strategies. New results are presented regarding the application of both the proposed method and the state-of-the-art approaches for various traffic control strategies.

3. Proposed Method

This section presents a method, which allows us to detect malicious data in urban vehicular networks. This method is intended for applications in urban road networks with signalized intersections. Main elements of the vehicular network are vehicles and road-side units. The vehicles periodically report their locations to road side units. The objective of the proposed approach is to recognize and filter out false location data that can be injected into the vehicular network by malicious vehicles. To this end a credibility score is determined for each vehicle. The credibility scores are updated by taking into account deviations from expected vehicle positions and inconsistencies between data from neighbouring vehicles. According to this method, the location data reported by vehicles with credibility scores below a predetermined threshold are categorized as malicious. Details of the above-mentioned operations are discussed in the following subsections.

3.1. Predicting Vehicle Position with Use of Cell Interval Traffic Model

The proposed method uses a road traffic model to estimate possible positions of vehicle. This model combines traffic cellular automata with interval representation of uncertain vehicle position and speed. The cellular automata approach involves discretization of time and space, as illustrated in Figure 1. According to this approach, the positions of vehicles along traffic lane are approximated by cells, i.e., non-overlapping lane segments of equal length. Length of one cell corresponds to average distance occupied by stopped vehicle in a completely congested road (7.5 m). In this study the cell length of 7.5 m was chosen based on literature [22,23]. The vehicle positions are updated by the traffic cellular automaton in steps of 1 second, that roughly estimate reaction time of average driver.

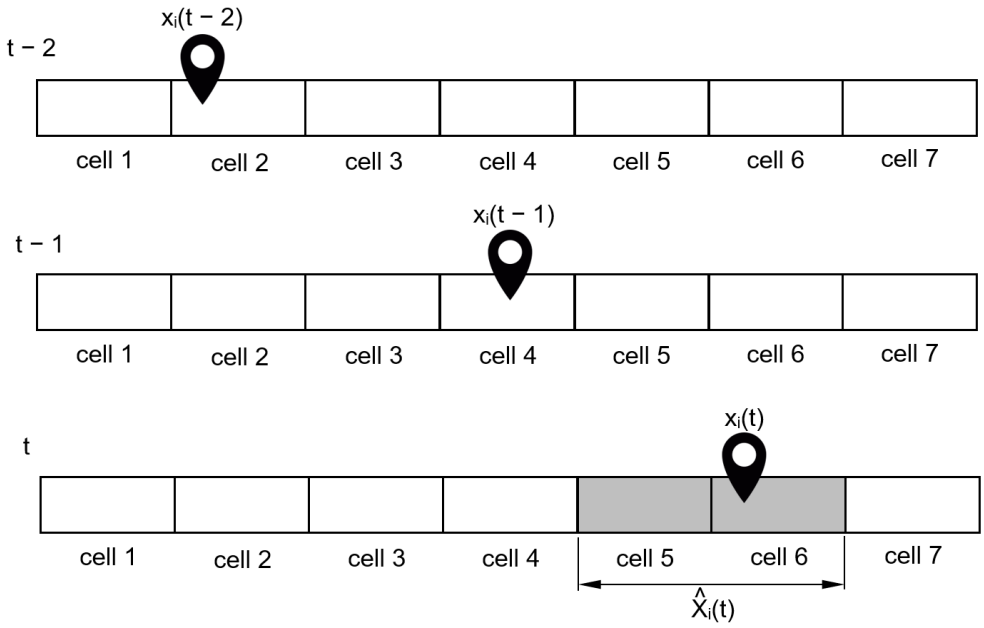

Figure 1. Positions of vehicle described by cell interval model for single traffic lane.

Positions reported by vehicles via the vehicular network are transformed to indices of the cells. For instance, Figure 1 shows the positions communicated by i-th vehicle for three successive time steps $(x_i(t-2), x_i(t-1), x_i(t))$. These positions are expressed in terms of cells as $c_i(t-2) = 2, c_i(t-1) = 4, c_i(t-2) = 6$, which means that according to the information delivered from i-th vehicle, this vehicle was in cell 2 at time step $t-2$, in cell 4 at time step $t-1$, and currently is in cell 6 (t denotes current time step).

In order to evaluate credibility of the above-mentioned information from vehicular network, the possible current locations of i-th vehicle are determined by using the cell interval traffic model. In other words, the traffic model allows us to find an interval of cells $\left[\hat{c}_i^-(t), \hat{c}_i^+(t)\right]$ that can be occupied by the considered vehicle at time step t. This cell interval is determined as follows. Firstly, the speed of i-th vehicle is evaluated for previous time step:

$$v_i(t-1) = c_i(t-1) - c_i(t-2). \tag{1}$$

Secondly, an update rule of the traffic cellular automaton is used to find interval $\left[\hat{v}_i^-(t), \hat{v}_i^+(t)\right]$, which describes the range of possible speed for current time step:

$$\begin{aligned} \hat{v}_i^-(t) &= \min\{v_i(t-1)+1, g_i(t-1), v_{\max}^-\}, \\ \hat{v}_i^+(t) &= \min\{v_i(t-1)+1, g_i(t-1), v_{\max}^+\}. \end{aligned} \tag{2}$$

It should be noted that the vehicle speed in this model is expressed in cells per time step (1 s). The maximum speed of vehicle is represented by the interval $[v_{\max}^-, v_{\max}^+]$, which has to reflect possible free flow velocities in the considered traffic stream.The quantity, denoted by $g_i(t-1)$, corresponds to number of free cells between i-th vehicle and the nearest obstacle ahead of it:

$$g_i(t-1) = co_i(t-1) - c_i(t-1) - 1, \tag{3}$$

where $co_i(t-1)$ is index of a cell of the nearest obstacle (another vehicle or red light).

The update rule of cellular automaton (Equation (2)) takes into account 3 assumptions that are necessary to reproduce the basic features of real traffic flow. The first assumption is that drivers have a tendency to drive as fast as possible. Thus, the drivers try to accelerate, i.e., increase the vehicle speed by 1 at each time step. The second assumption is necessity to avoid collisions. It says that the speed (in cells per time step) cannot exceed the number of free cells in front of the vehicle (g_i). The third assumption relates to the fact that drivers have to obey a speed limit ($v_{\max}$).

Finally, the interval of cells that can be currently occupied by i-th vehicle is determined using formulas:

$$\begin{aligned} \hat{c}_i^-(t) &= c_i(t-1) + \hat{v}_i^-(t), \\ \hat{c}_i^+(t) &= c_i(t-1) + \hat{v}_i^+(t). \end{aligned} \tag{4}$$

These formulas assume that the vehicle is moved according to the new velocity determined by Equation (2).

In Figure 1, the above defined cell interval is indicated by gray color, i.e., $\hat{c}_i^-(t) = 5$ and $\hat{c}_i^+(t) = 6$. This integer valued interval is transformed to a real valued interval $\hat{X}_i(t) = \left[\hat{x}_i^-(t), \hat{x}_i^+(t)\right]$, which describes the uncertain expected position in the domain of the continuous coordinate. As shown in Figure 1, the endpoint $\hat{x}_i^-(t)$ corresponds to localisation of the left boundary of cell $\hat{c}_i^-(t)$ and the endpoint $\hat{x}_i^+(t)$, that corresponds to the right boundary of cell $\hat{c}_i^+(t)$. Based on this observation, the real valued interval $\hat{X}_i(t)$ is easily determined as the locations of cell boundaries, which are set during calibration of the traffic model, when each traffic lane is divided into the 7.5 m long cells.

The above discussed cell interval model is also used for determining possible vehicle positions in multilane traffic. In case of vehicle i, which is moving in traffic lane l, its possible positions can be evaluated for adjacent lanes ($l+1$ and/or $l-1$). To this end it is verified if cell $c_i(t-1)$ is empty in the adjacent lane. If this condition is met, the cell interval $\left[\hat{c}_i^-(t), \hat{c}_i^+(t)\right]$ for the adjacent lane is determined in accordance with Equations (1)–(4), assuming that the vehicle is present in the target (adjacent) lane at time step $t-1$.

Figure 2 shows an example of determining possible vehicle positions for three lane road. It was assumed that direction of the traffic flow corresponds to the increasing cell indices. The cells with black rectangles show positions of vehicles at time step $t-1$, while the cells filled with particular colours correspond to the possible vehicle positions at time step t. For instance, in case of vehicle 1, which occupies cell 1 in lane 2 at time step $t-1$, the possible positions for time step t are represented by interval $\left[\hat{c}_1^-(t), \hat{c}_1^+(t)\right] = [2,3]$ for all three lanes. In case of vehicle 2 the interval $\left[\hat{c}_1^-(t), \hat{c}_1^+(t)\right]$ equals [5, 5] for lane 1, [5, 6] for lane 2, and is empty for lane 3. It should be noted that the expected vehicle positions in Figure 2 were determined for $v_i(t-1) = 2, i = 1 \ldots 6$ and $[v_{\max}^-, v_{\max}^+] = [1,2]$. The expected positions for different vehicles may overlap. For instance, the model in Figure 2 assumes that at time step t, cell 10 in lane 2 can be occupied by vehicle 4, 5 or 6. The overlapping expected positions are not shown in Figure 2 for clarity of presentation. For the same reason, the possible positions of vehicles 4 and 6 are not indicated by colours. In case of vehicle 4, the possible positions correspond to interval $[9,10]$ for lanes 2 and 3. For vehicle 6 the possible positions in traffic lanes 1 and 2 are given by interval $[10,11]$.

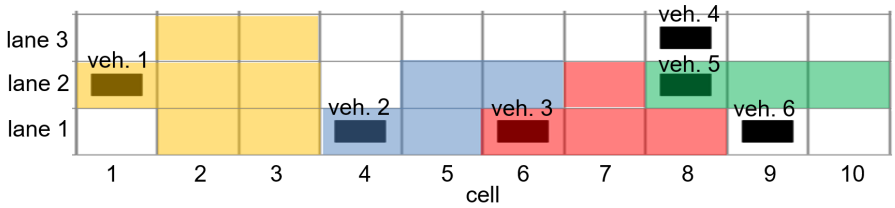

Figure 2. Possible vehicles positions determined for multilane traffic.

3.2. Evaluation of Credibility Scores

The possible vehicles positions, predicted by the cell interval traffic model, are used to evaluate their credibility scores. Details of the credibility evaluation are presented by the pseudo-code of Algorithm 1. Input data of the credibility evaluation algorithm include the traffic lane identifiers (l_i) and positions reported by vehicles for the last three time steps. It should be noted that symbol m in Algorithm 1 denotes the number of vehicles currently registered by a roadside unit. Moreover, the input data includes previous vehicle credibility scores (cs_i) that have to be updated. The credibility score of newly registered vehicles is set to zero. In line 5 of Algorithm 1, the traffic model discussed in Section 3.1 is used to find possible vehicle positions. It means that the interval $\hat{X}_i(t)$, which describes the uncertain expected position of a vehicle, is calculated based on Equations (1)–(4). If the current position reported by a vehicle $x_i(t)$ coincides with the predicted possible position, then the vehicle credibility score is increased by α. In opposite situation, the score is decreased by α (see lines 6–10 of Algorithm 1).

Subsequently, the credibility score algorithm takes into account queuing of vehicles. When a traffic queue is formed, the fact that a vehicle with positive credibility score stops in cell $c-1$ confirms presence of another vehicle ahead in cell c. In such situation, the credibility score of the vehicle in cell c is increased by α (see lines 11–17 of Algorithm 1). It should be noted that the condition "vehicle i is stopped next to j" in the pseudo code of Algorithm 1 is satisfied when the vehicles i and j are stopped in cells $c-1$ and c, respectively.

The proposed algorithm also detects inconsistencies in the position data delivered from vehicles that correspond to unrealistic traffic situations. This operation is based on the concept of time-space trajectories that can be presented in form of a diagram, where time is denoted on the horizontal axis and distance from a reference point on the vertical axis. Figure 3 shows an example of time-space trajectories for three vehicles. In this example vehicle 1 decelerates and two remaining vehicles are travelling with a constant speed. Note that the three vehicles move in one traffic lane. Since vehicles cannot overtake on single lane, the intersecting trajectories (time step 4, cell 8 and time step 8, cell 9) should be considered as unrealistic

situations. The intersecting time-space trajectories of vehicles i and j at time step t are recognized using Algorithm 2. This algorithm checks if vehicles are in the same traffic lane (line 2) and verifies if order of vehicles changes, by taking into account cells occupied by the vehicles (lines 2 and 4). When time-space trajectory of a first vehicle is intersected by trajectory of a second vehicle, which has positive credibility score, then credibility score of the first vehicle is decreased by β (see lines 18–25 of Algorithm 1).

Algorithm 1 Evaluation of credibility scores

Input: $T = \{\langle x_i(t-\tau), l_i(t-\tau)\rangle : i = 1 \dots m, \tau = 0 \dots 2\}, cs_1 \dots cs_m$
Output: updated $cs_1 \dots cs_m$
1: **for** $i = 1 \dots m$ **do**
2: **if** vehicle i is new **then**
3: $cs_i := 0$
4: **end if**
5: use traffic model to predict $\hat{X}_i(t)$ for lane $l_i(t)$ based on T
6: **if** $x_i(t) \in \hat{X}_i(t)$ **then**
7: $z_i = \alpha$
8: **else**
9: $z_i = -\alpha$
10: **end if**
11: **for** $j = 1 \dots i-1$ **do**
12: **if** vehicle i is stopped next to j and $cs_i > 0$ **then**
13: $z_j := z_j + \alpha$
14: **end if**
15: **if** vehicle j is stopped next to i and $cs_j > 0$ **then**
16: $z_i := z_i + \alpha$
17: **end if**
18: **if** time-space trajectories of vehicles i and j intersects **then**
19: **if** $cs_i > 0$ **then**
20: $z_j := z_j - \beta$
21: **end if**
22: **if** $cs_j > 0$ **then**
23: $z_i := z_i - \beta$
24: **end if**
25: **end if**
26: **end for**
27: **end for**
28: **for** $i = 1 \dots m$ **do**
29: $cs_i := \min(\max(cs_i + z_i, cs_{\min}), cs_{\max})$
30: **end for**

Algorithm 2 Detection of intersecting time-space trajectories.

1: **if** $l_i(t) = l_j(t)$ **then**
2: **if** $c_i(t) \geq c_j(t)$ and $c_i(t-1) < c_j(t-1)$ **then**
3: **return** true
4: **end if**
5: **if** $c_j(t) \geq c_i(t)$ and $c_j(t-1) < c_i(t-1)$ **then**
6: **return** true
7: **end if**
8: **end if**
9: **return** false

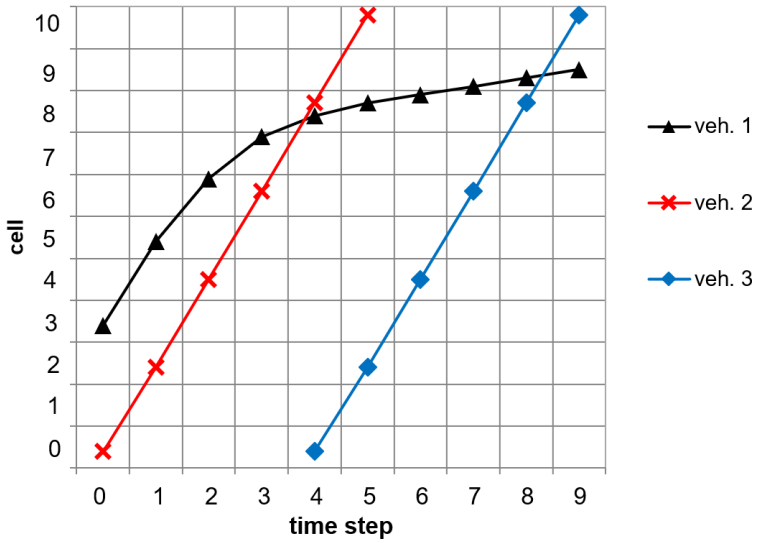

Figure 3. Intersecting time-space trajectories.

Algorithm 1 uses two parameters (α and β) for updating the credibility scores. The two different parameters were introduced to distinguish between the updates that can occur more frequently (related to possible positions and stopping vehicles) from those that are less frequent (intersecting time-space trajectories). The changes of credibility score should be smaller in case of the more frequent updates and bigger for the potentially rare events. The examples presented in this Section (Figures 4 and 5) assume that $\alpha = 0.1$ and $\beta = 1$ for illustrative purposes. In case of the experiments reported in Section 4, α was set to 0.2 and β to 1. These values were selected as they provided the best results of malicious data detection during preliminary tests.

It should be also noted that the range of credibility score values is limited by the maximum $cs_{\max}$ and minimum $cs_{\min}$ (see line 28 of Algorithm 1). Based on preliminary experiments, it was assumed that $cs_{\max} = 30$ and $cs_{\min} = -30$.

Examples of credibility score evaluation are presented in Figures 4 and 5. These examples take into account vehicles in single traffic lane at a signalized intercession. In the first scenario (Figure 4) a traffic light, at the intersection, for the considered vehicles is red, while in the second scenario (Figure 5) the traffic signal is green. It should be noted that Figures 4a and 5a show vehicle positions in time-space diagrams. The black data points in time-space diagrams correspond to positions reported by vehicles ($x_i(t)$). Additionally, the expected vehicle positions ($\hat{X}_i(t)$) are depicted by cells filled with colours. Figures 4b and 5b depict the credibility scores evaluated for particular vehicles in accordance with Algorithm 1 ($\alpha = 0.1$, $\beta = 1$). In these examples it was assumed that the maximum speed interval $[v^-_{\max}, v^+_{\max}]$ equals [1, 2] (in cells per time step).

In case of the first scenario (Figure 4) all vehicles should stop in a queue due to the red signal (which is not shown in the chart). However, a selfish vehicle reports additional false vehicles (veh. 3 and veh. 4) to get the green light faster. The true vehicles (veh. 1 and veh. 2) stop at the end of the queue. The false (malicious) time-space trajectories of vehicles 3 and 4 cross the trajectories of stopped vehicles. It should be noted that the attacking selfish vehicle does not know the positions of true vehicles thus, it cannot generate realistic false trajectories. As a result, the credibility scores for vehicles 3 and 4 fall below zero at time steps 7 and 9, which means that these vehicles are correctly recognized as malicious by the proposed algorithm.

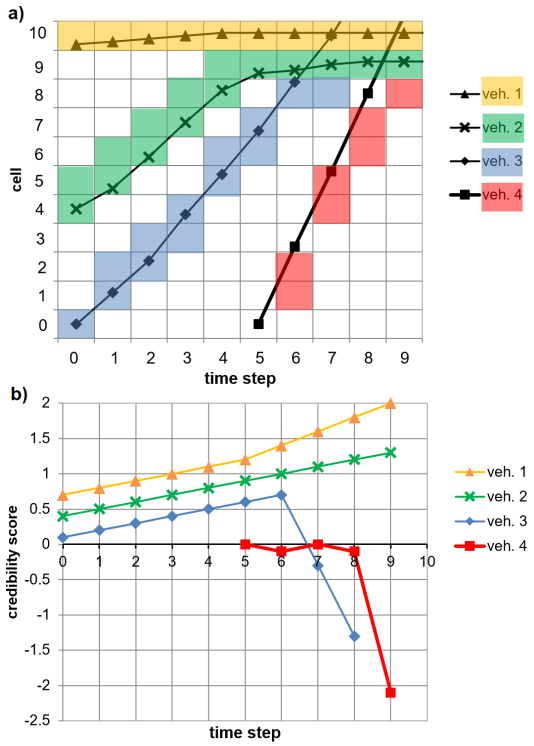

Figure 4. Vehicles approaching red traffic light: (**a**) time-space trajectories (**b**) credibility scores.

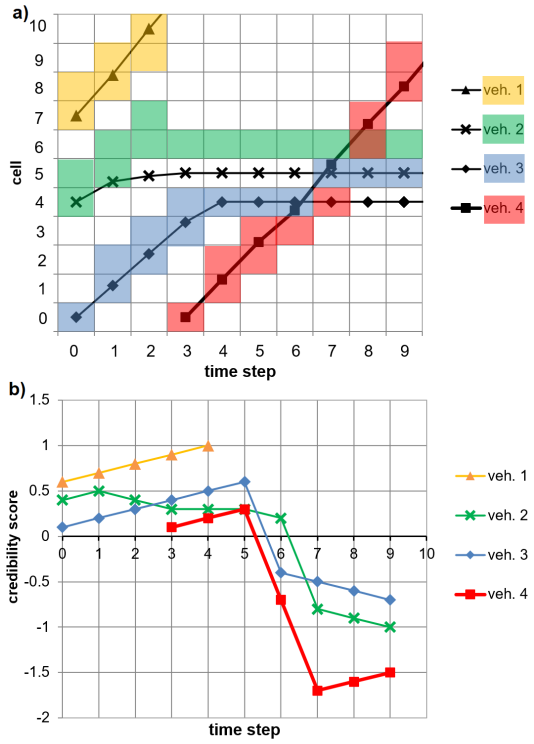

Figure 5. Vehicles approaching green traffic light: (**a**) time-space trajectories (**b**) credibility scores.

The red traffic light is not present in the second scenario (Figure 5) thus, all vehicles should pass the considered road section without stopping (similarly to vehicle 1). However, in this example vehicles with malfunctioning sensors are present (veh. 2 and veh. 3). These vehicles report the malicious time-space trajectories, which suggest that they stop in cells 4 and 5. The trajectories are correctly detected to be false when vehicle 4 passes cells 4 and 5. It should be also noted that the credibility score drops below zero also for the true vehicle (veh. 4), however this score has a positive trend in time steps 7–9 and the vehicle will be categorized as true again after several time steps. Since vehicles 2 and 3 are recognized as malicious, the credibility score will not decrease for subsequent vehicles passing cells 4 and 5.

4. Experiments

The proposed method of malicious data detection was experimentally evaluated in a real-world simulation scenario of a road network in Chicago. Figure 6 presents a map of this road network. The simulated roads and intersections are marked with yellow colour. The roads in this scenario have one or two traffic lanes for each direction. Each of the sixteen simulated intersections is controlled by traffic lights. Examples of the simulated four-way intersections are presented in Figure 7. The simulations were conducted in SUMO (Simulation of Urban MObility) [24]. Maximum speed of the simulated vehicles was 50 km/h.

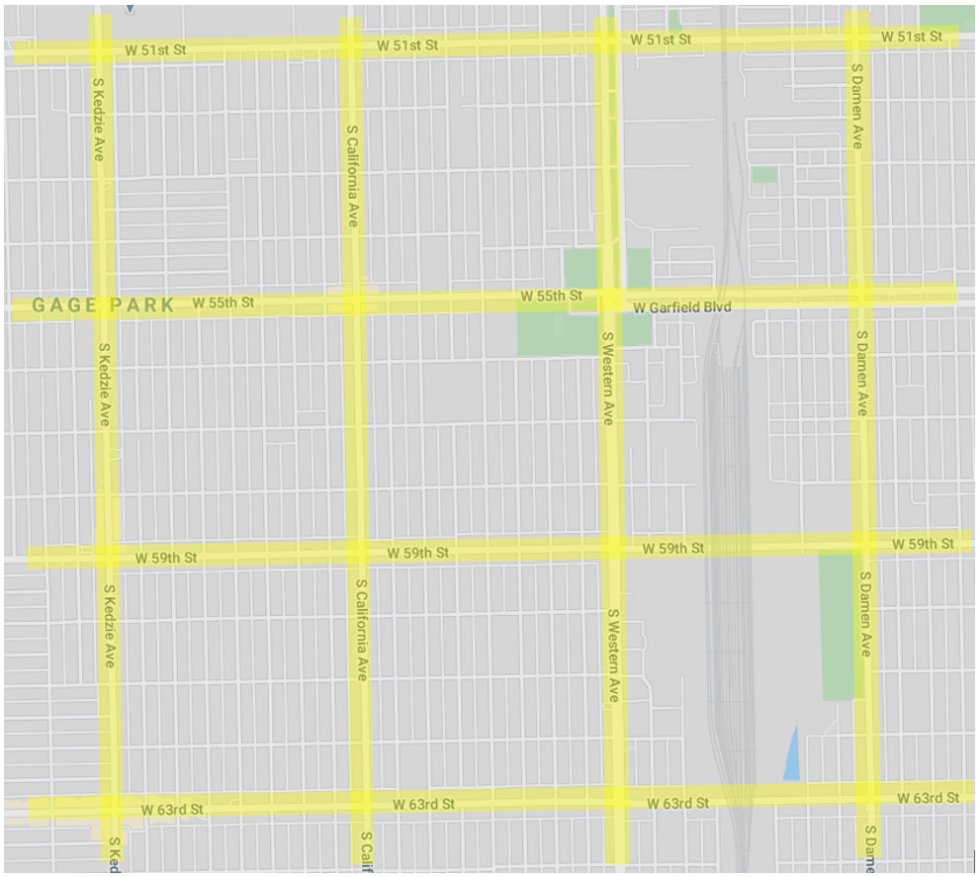

Figure 6. Simulated urban road network with signalized intersections (based on Google maps).

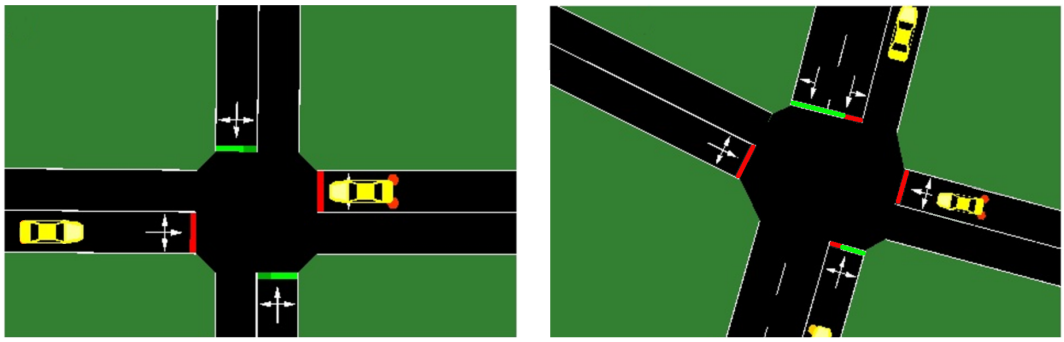

Figure 7. Examples of simulated intersections.

In the simulation experiments it was assumed that all vehicles report their positions, via vehicular communication network, to road side units installed at intersections. Each intersection in this model has a dedicated road side unit, which collects data from vehicles in neighbouring roads and controls the traffic signals using an adaptive self-organizing strategy. Malicious data were simulated by reporting false vehicle positions to road side units. Two types of the malicious data were generated. The first type takes into account the vehicles that performs Sybil attack. A vehicle of this kind reports multiple false vehicle positions that corresponds to its own locations registered at previous time steps. The second type of malicious data is obtained by simulating false vehicles that move with random velocities selected from interval [0, 60] km/h.

During experiments the proposed approach to malicious data detection was compared with two representative methods from the related literature. The first method [13] assumes that vehicles are equipped with sensors that enables localization of the objects in its surrounding. The location of a vehicle, which sends a message, has to be verified by the neighbouring vehicles. This method is denoted as "Method 1" in charts. The second method [17] uses fundamental traffic flow theory for detection of anomalous speed and headway between vehicles. On this basis the trustworthiness of messages broadcasted by vehicles is evaluated. This method is denoted in charts as "Method 2".

For the proposed approach, the parameters α and β were equal to 0.2 and 1, respectively. These values were selected based on preliminary results. The traffic cellular automaton was used with maximum speed interval [1, 2] cells per time step.

The quality of malicious data detection was evaluated, for the compared methods, by taking into account balanced accuracy and sensitivity. In most cases the amount of malicious and true data differs significantly. To deal with the imbalanced data set, the balanced accuracy BA is considered as the main metric:

$$BA = 0.5 \cdot \left(\frac{TP}{P} + \frac{TN}{N} \right), \tag{5}$$

where P and N denote the numbers of malicious and true data, respectively. TP is the number of correctly detected malicious data, and TN is the number of correctly recognized true data. Additionally, the sensitivity, defined as TP/P is analysed. It should be noted that the data are transmitted by vehicles and categorized as malicious or not in one second intervals (time steps).

The experiments were conducted for different intensities of the true and malicious vehicles. The intensity of true vehicles was changed between 0.02 and 0.14 vehicles per second, which allows us to analyse both low and high traffic scenarios (including congestion). For false (malicious) vehicles the intensity varies from 0.02 to 0.1. It should be noted that the intensity is understood as average number of vehicles which enter each simulated road in one second period. One-hour simulation was executed 10 times for each combination of the intensities. Summarized results, for all simulations, are presented

in Figure 8. These results show that the proposed method detects the malicious data with the highest balanced accuracy and the highest sensitivity, outperforming the other compared methods. The worst results were obtained for Method 2, which takes into account average headway and speed of vehicles and compares these parameters with theoretical values determined on the basis of traffic theory. This approach does not work well for roads with signalized intersections, where high spatial and temporal variations of speed and headway are usually observed.

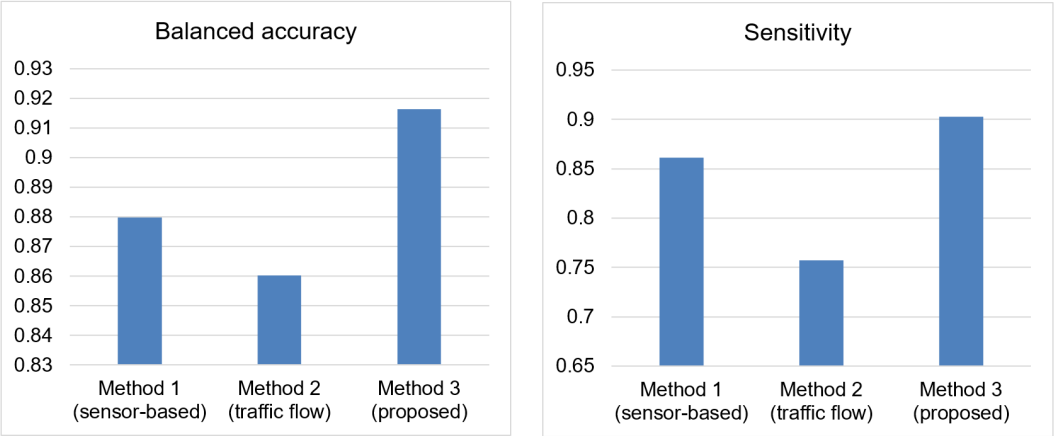

Figure 8. Accuracy and sensitivity of malicious data detection methods.

The impact of traffic intensity on accuracy of malicious data detection is illustrated in Figure 9. The traffic intensity in this chart corresponds to intensity of true vehicles. For all intensities of true vehicles, the proposed method achieved the highest balanced accuracy. In case of all compared methods, the accuracy increases with traffic intensity. However, the increase is faster for Method 1 and Method 3 (proposed). The reason behind this observation is that in case of Method 1 the false vehicles are detected by sensors of the true vehicles, thus the accuracy of malicious data detection is improved when the number of true vehicles is higher. Similarly, using the proposed approach, it is easier to detect the false vehicles when the number of true vehicles is high and inconsistencies between true and malicious data reports (e.g., intersecting time-space trajectories) occur more frequent.

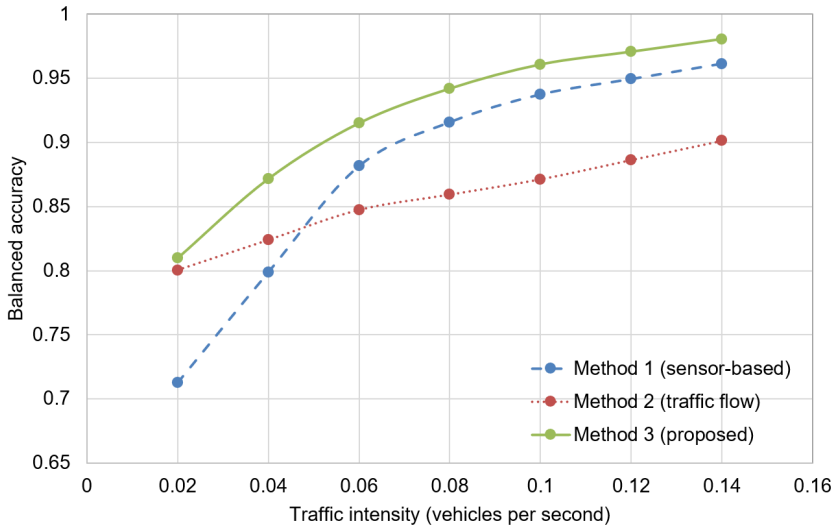

Figure 9. Accuracy of malicious data detection for different traffic intensities.

Accuracy of the compared methods, for different intensities of malicious vehicles, is analysed in Figure 10. The accuracy of malicious data detection slightly decreases with increase of intensity of false vehicles. In case of method 1 this effect is not visible in Figure 10 as the differences are very small. Again, the best results were obtained for the new proposed method.

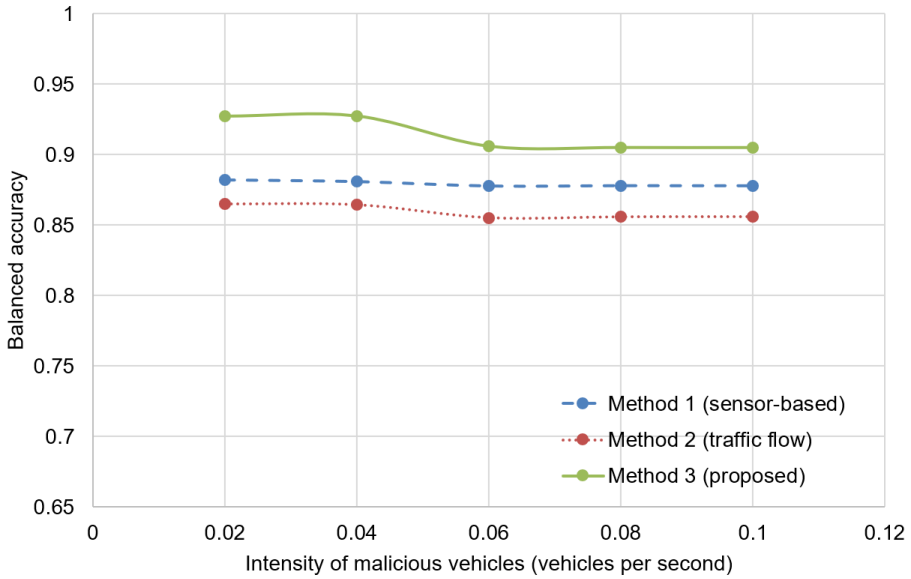

Figure 10. Accuracy of malicious data detection for different intensities of malicious vehicles.

Figure 11 compares the accuracy of malicious data detection for three cases. In the first case (left part of the chart) the number of simulated true vehicles (T) was higher than the number of malicious vehicles (M). The second case concerns simulations for which the numbers of malicious and true vehicles were equal (T = M). Finally, the right part of the chart shows the results obtained for situations when the number of malicious vehicles was higher (T < M). In this case (for T < M) the lowest accuracy was observed for all compared methods. The results in Figure 11 shows that the proposed method outperforms the state-of-the-art approaches for each analysed case, even if the false vehicles are prevailing.

The last part of the experimental results, discussed in this section, concerns the performance of traffic signal control in different scenarios of malicious data detection. As it was already mentioned, the traffic signals in the simulated road network are controlled by using self-organizing control strategies. The self-organizing signal control scheme enables a decentralized optimization and global coordination of the traffic signals at many intersections. According to the self-organizing strategy, traffic signals at each intersection in the road network are controlled independently on the basis of real-time traffic data transmitted from vehicles present in road segments connected to the given intersection. Four different self-organizing strategies were considered in this study: back-pressure strategy (BP) [25], strategy proposed by Lämmer and Helbing [26], the self-organizing system based on predictive interval microscopic model (SOS) [2], and the control strategy based on neuroevolution (Neuro) [27].

Performance of traffic signal control was compared for the above-mentioned strategies by taking into account total vehicle delay (Figure 12) and average speed (Figure 13). It should be noted here that input data of the above-mentioned control algorithms include locations of the vehicles with credibility score above zero. As shown in Figures 12 and 13, the proposed method allows us to obtain comparable results with the scenarios without malicious data. For each control strategy, the proposed method outperformed other methods. Additionally, it is worth to note that each analysed control strategy utilizes the input data in a different way (e.g., in BP strategy the total number of vehicles in particular lanes is taken into account,

while for the Neuro strategy positions of individual are also important). Thus, it can be concluded that the proposed method is useful for simple as well as for complex control strategies. Finally, the best results can be achieved using the control strategy based on neuroevolution (Neuro).

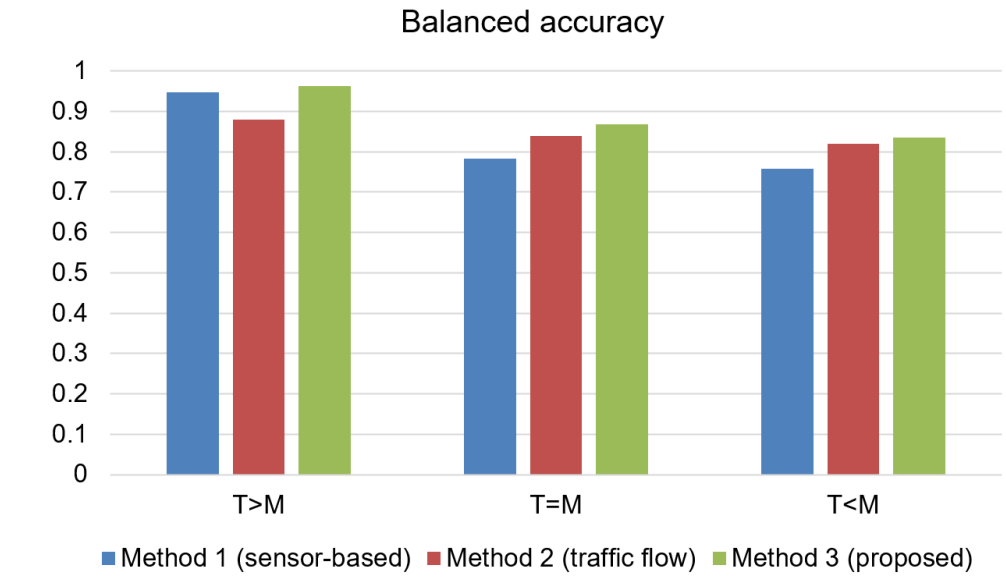

Figure 11. Accuracy of malicious data detection methods for different share of malicious vehicles.

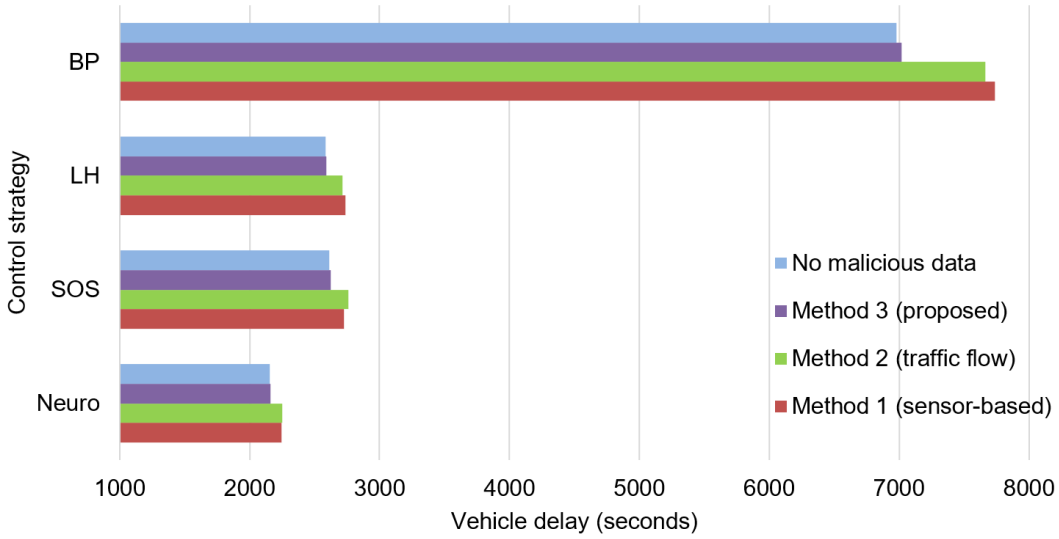

Figure 12. Delay of vehicles for different signal control strategies with and without detection of malicious data.

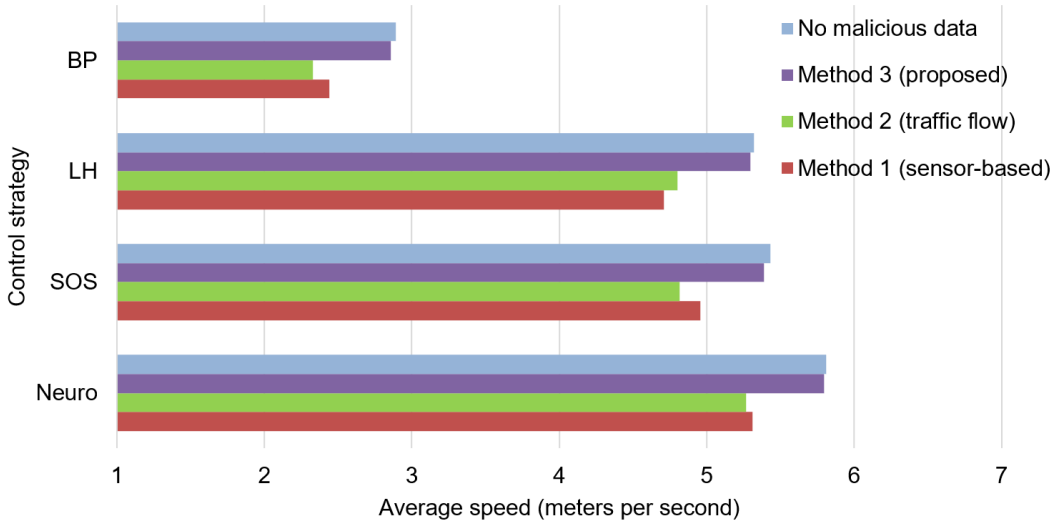

Figure 13. Average speed of vehicles for different signal control strategies with and without detection of malicious data.

5. Conclusions

Presence of malicious data impede effective use of information collected from vehicles via urban vehicular networks. The malicious data have to be eliminated in real time if the vehicular network is a source of information for the control systems, that manage traffic signals at intersections. In order to filter out the malicious data, a credibility score algorithm was proposed, which determines possible locations of individual vehicles by using a microscopic traffic model. This algorithm is computationally effective and does not require any additional exchange of data between vehicles nor installation of dedicated sensors. Thus, the introduced method is easy for implementation in smart cites. The proposed algorithm is suitable for applications in intelligent transportation systems that collect fine-grained information, regarding individual vehicles, for control purposes. The role of this algorithm is to improve safety and effectiveness of the above-mentioned applications by rejecting false information.

Extensive simulation experiments were conducted to verify effectiveness of the proposed approach in realistic scenarios. The experiments considered an urban road network with traffic signals at many intersections. The traffic signals were controlled by using various decentralized self-organizing strategies. The experimental results show that the proposed method detects the malicious data with higher accuracy than the considered state-of-the-art methods. The improved accuracy of detecting malicious data has enabled mitigation of their negative impact on the performance of traffic signal control.

Main limitations of the proposed approach are related to the necessity of a detailed parameter calibration. Especially the parameters of the traffic model have to be carefully calibrated. Otherwise, the model incorrectly predicts the possible driver behaviour. Another drawback is a decrease of accuracy for high fractions of malicious data.

Further research directions include extending the method to enable automatic adaptation of the traffic model parameters based on the collected data. Another interesting topic for future research is related to possible modifications of the proposed approach for traffic monitoring systems based on edge computing and applications in autonomous vehicles.

Author Contributions: Conceptualization, B.P.; Data curation, B.P. and M.B.; Formal analysis, B.P.; Funding acquisition, B.P.; Investigation, B.P. and M.B.; Methodology, B.P.; Project administration, B.P.; Resources, B.P. and M.C.; Software, B.P. and M.B.; Supervision, B.P.; Validation, B.P. and M.C.; Visualization B.P. and M.B.; Writing—original draft B.P.; Writing—review & editing, B.P., M.B., and M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The research reported in this paper was partially inspired by the results obtained under Grant No. LIDER/18/0064/L-7/15/NCBR/2016.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bermad, N.; Zemmoudj, S.; Omar, M. Context-aware negotiation, reputation and priority traffic light management protocols for VANET-based smart cities. *Telecommun. Syst.* **2019**, *72*, 131–153. [CrossRef]
2. Płaczek, B. A self-organizing system for urban traffic control based on predictive interval microscopic model. *Eng. Appl. Artif. Intell.* **2014**, *34*, 75–84. [CrossRef]
3. Płaczek, B.; Bernas, M. Detection of malicious data in vehicular ad hoc networks for traffic signal control applications. In *International Conference on Computer Networks*; Springer: Cham, Switzerland, 2016; pp. 72–82.
4. Gu, K.; Dong, X.; Jia, W. Malicious Node Detection Scheme Based on Correlation of Data and Network Topology in Fog Computing-based VANETs. *IEEE Trans. Cloud Comput.* **2020**. [CrossRef]
5. Rezgui, J.; Doucet, C. Detection of malicious vehicles with demerit and reward level system. In Proceedings of the 2017 International Symposium on Networks, Computers and Communications (ISNCC), Marrakech, Morocco, 16–18 May 2017; pp. 1–6.
6. Yang, Y.; Ou, D.; Xue, L.; Dong, D. Infrastructure-based Detection Scheme of Malicious Vehicles for Urban Vehicular Network (No. 17-05475). In Proceedings of the Transportation Research Board 96th Annual Meeting, Washington, DC, USA, 8–12 January 2017.
7. Arshad, M.; Ullah, Z.; Ahmad, N.; Khalid, M.; Criuckshank, H.; Cao, Y. A survey of local/cooperative-based malicious information detection techniques in VANETs. *Eurasip J. Wirel. Commun. Netw.* **2018**, *2018*, 62. [CrossRef]
8. Hasrouny, H.; Samhat, A.E.; Bassil, C.; Laouiti, A. VANet security challenges and solutions: A survey. *Veh. Commun.* **2017**, *7*, 7–20. [CrossRef]
9. Ghosh, M.; Varghese, A.; Gupta, A.; Kherani, A.A.; Muthaiah, S.N. Detecting misbehaviors in VANET with integrated root-cause analysis. *Ad Hoc Netw.* **2010**, *8*, 778–790. [CrossRef]
10. Vulimiri, A.; Gupta, A.; Roy, P.; Muthaiah, S.N.; Kherani, A.A. Application of secondary information for misbehavior detection in VANETs. In *International Conference on Research in Networking*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 385–396.
11. Sun, M.; Li, M.; Gerdes, R. A data trust framework for VANETs enabling false data detection and secure vehicle tracking. In Proceedings of the 2017 IEEE Conference on Communications and Network Security (CNS), Las Vegas, NV, USA, 9–11 October 2017; pp. 1–9.
12. Arshad, M.; Ullah, Z.; Khalid, M.; Ahmad, N.; Khalid, W.; Shahwar, D.; Cao, Y. Beacon trust management system and fake data detection in vehicular ad-hoc networks. *IET Intell. Transp. Syst.* **2018**, *13*, 780–788. [CrossRef]
13. Lim, K.; Tuladhar, K.M.; Kim, H. Detecting location spoofing using ADAS sensors in VANETs. In Proceedings of the 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 11–14 January 2019; pp. 1–4.
14. Azuma, S.; Tsukada, M.; Nomura, T.; Sato, K. A Method of Detecting Camouflage Data with Mutual Vehicle Position Monitoring. In Proceedings of the Sixth International Conference on Advancesin Vehicular Systems, Technologies and Applications (VEHICULAR 2017), Nice, France, 6–9 June 2017.
15. Van der Heijden, R.; Dietzel, S.; Kargl, F. Misbehavior detection in vehicular ad-hoc networks. In Proceedings of the 1st GI/ITG KuVS Fachgespräch Inter-Vehicle Communication (FG-IVC 2013), Tyrol, Austria, 21–22 February 2013.

16. Ghaleb, F.A.; Maarof, M.A.; Zainal, A.; Rassam, M.A.; Saeed, F.; Alsaedi, M. Context-aware data-centric misbehaviour detection scheme for vehicular ad hoc networks using sequential analysis of the temporal and spatial correlation of the consistency between the cooperative awareness messages. *Veh. Commun.* **2019**, *20*, 100186. [CrossRef]
17. Ranaweera, M.; Seneviratne, A.; Rey, D.; Saberi, M.; Dixit, V.V. Anomalous data detection in vehicular networks using traffic flow theory. In Proceedings of the 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall), Honolulu, HI, USA, 22–25 September 2019; pp. 1–5.
18. Van der Heijden, R.W.; Dietzel, S.; Leinmüller, T.; Kargl, F. Survey on misbehavior detection in cooperative intelligent transportation systems. *IEEE Commun. Surv. Tutor.* **2018**, *21*, 779–811. [CrossRef]
19. Yavvari, C.; Duric, Z.; Wijesekera, D. Vehicular dynamics based plausibility checking. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–8.
20. Kerrache, C.A.; Lakas, A.; Lagraa, N.; Barka, E. UAV-assisted technique for the detection of malicious and selfish nodes in VANETs. *Veh. Commun.* **2018**, *11*, 1–11. [CrossRef]
21. Lu, Z.; Wang, Q.; Qu, G.; Liu, Z. Bars: A blockchain-based anonymous reputation system for trust management in vanets. In Proceedings of the 2018 17th IEEE International Conference On Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, USA, 1–3 August 2018; pp. 98–103.
22. Janczykowski, M.; Turek, W.; Malawski, M.; Byrski, A. Large-scale urban traffic simulation with Scala and high-performance computing system. *J. Comput. Sci.* **2019**, *35*, 91–101. [CrossRef]
23. Ruan, X.; Zhou, J.; Tu, H.; Jin, Z.; Shi, X. An improved cellular automaton with axis information for microscopic traffic simulation. *Transp. Res. Part C Emerg. Technol.* **2017**, *78*, 63–77. [CrossRef]
24. Krajzewicz, D.; Erdmann, J.; Behrisch, M.; Bieker, L. Recent development and applications of SUMO-Simulation of Urban MObility. *Int. J. Adv. Syst. Meas.* **2013**, *5*, 4.
25. Zaidi, A.A.; Kulcsár, B.; Wymeersch, H. Back-pressure traffic signal control with fixed and adaptive routing for urban vehicular networks. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2134–2143. [CrossRef]
26. Lämmer, S.; Helbing, D. Self-control of traffic lights and vehicle flows in urban road networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P04019. [CrossRef]
27. Bernas, M.; Płaczek, B.; Smyła, J. A neuroevolutionary approach to controlling traffic signals based on data from sensor network. *Sensors* **2019**, *19*, 1776. [CrossRef] [PubMed]