Distributed Classification of Traffic Anomalies Using Microscopic Traffic Variables

Suttipong Thajchayapong, Edgar S. Garcia-Treviño, and Javier A. Barria, *Member, IEEE*

***Abstract*—This paper proposes a novel anomaly classification algorithm that can be deployed in a distributed manner and utilizes microscopic traffic variables shared by neighboring vehicles to detect and classify traffic anomalies under different traffic conditions. The algorithm, which incorporates multiresolution concepts, is based on the likelihood estimation of a neural network output and a bisection-based decision threshold. We show that, when applied to real-world traffic scenarios, the proposed algorithm can detect all the traffic anomalies of the reference test data set; this result represents a significant improvement over our previously proposed algorithm. We also show that the proposed algorithm can effectively detect and classify traffic anomalies even when the following two cases occur: 1) the microscopic traffic variables are available from only a fraction of the vehicle population, and 2) some microscopic traffic variables are lost due to degradation in vehicle-to-vehicle (V2V) or vehicle-to-infrastructure communications (V2I).**

***Index Terms*—Distributed traffic monitoring, freeway segments, incident precursors, microscopic traffic variables, traffic anomalies detection, vehicle to infrastructure (V2I), vehicle to vehicle (V2V).**

I. Introduction

THE FIRST step in detecting the possible emergence of traffic incidents is to identify *traffic anomalies*, which are defined in this paper as traffic patterns that do not conform to expected behavior [1]–[3]. Once the detected abnormal traffic pattern has been identified, information with regard to its potential impacts can be disseminated to individual drivers and traffic management centers, such as the Advanced Traveler Information System and Advanced Traffic Management System, so that the appropriate proactive strategies for minimizing the response times, clearing the roadways, and recovering traffic conditions back to normality can promptly be set in place.

Manuscript received April 15, 2011; revised September 30, 2011, January 19, 2012, April 9, 2012, and June 14, 2012; accepted September 10, 2012. Date of publication October 19, 2012; date of current version February 25, 2013. The Associate Editor for this paper was S. Tang.

S. Thajchayapong is with the National Electronic and Computer Technology Center, National Science and Technology Development Agency, Pathumthani 12120, Thailand.

E. S. Garcia-Treviño and J. A. Barria are with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, U.K. (e-mail: j.barria@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2012.2220964

Recent advances in vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) wireless communications have increased the potential of real-time measuring and processing of microscopic traffic variables from vehicles in a distributed manner. Distributed monitoring in this paper refers to the process by which traffic anomalies are detected and classified by local vehicles themselves and/or roadside infrastructure based on local traffic information on the road segment of interest, without the need to send traffic information to a traffic management center. We will refer to these local vehicles and/or roadside infrastructure as *nodes* in the remainder of this paper.

Recent studies have attempted to develop algorithms to detect traffic anomalies that lead to traffic disruptions, commonly referred to as *incident precursors* [4]–[6], an incident that is a nonrecurring major traffic disruption whose occurrence is usually unexpected and random [7]. However, because most of these algorithms are based on macroscopic traffic variables derived only from roadside infrastructure, such as loop detectors, their effectiveness largely depends on the relative location of the anomaly with respect to the loop detectors. If a disruption takes place far from the loop detector location, the anomaly may not be detected, or a long delay may occur before the anomaly is identified. On the other hand, microscopic traffic information that is obtained from vehicles has the advantages of coverage and sensitivity to support distributed anomaly detection schemes. The results presented in this paper and in previous findings [1], [8] strongly suggest that algorithms that can use available microscopic traffic variables for anomaly detection and classification are worth developing. In this paper, we present a novel algorithm that can detect traffic anomalies in a distributed manner.

To use microscopic traffic information in a distributed manner, an algorithm has to, at the very least, possess the following properties. First, the algorithm needs to be deployable on a local node, where there may be limited local storage and communication capabilities. Second, an anomaly detection and classification algorithm not only needs to detect different types of traffic anomalies but should also minimize the false-alarm rate (FAR) and hence reduce emergency response costs. Third, an early warning signal has to be available as fast as possible to give enough lead time to operations/management personnel to initiate appropriate emergency responses before the anomaly evolves into a major traffic disruption. Here, we note that previously proposed algorithms have been assessed on maximizing the detection rate (DR) but not properly tested on FAR and mean time to detection (MTTD) [4]. Last, the algorithm should be adaptable to different traffic conditions such as a low-flow low-speed condition during rush hours and a low-flow

high-speed condition during daytime periods. Most of the existing algorithms use a prespecified time window (5 min in [4] and 8, 3, and 2 min in [6]), which is chosen to maximize the difference between traffic variables under anomalous conditions and those under normal conditions. We have shown in [1] and [9] that the effectiveness of an anomaly detection algorithm of this type is highly dependent on the choice of the size of its time window. That is, longer time windows are needed to capture changes under low-flow high-speed conditions, and vice versa. Therefore, to be deployable, an anomaly detection algorithm based on microscopic traffic variables needs to identify different types of traffic anomalies under different traffic conditions.

In this paper, we propose a novel anomaly detection and classification algorithm that uses microscopic traffic variables and exhibits significant performance improvement over the algorithms presented in [1]. Unlike other previously reported approaches [4], [6], where a particular interval prior to incidents has to be selected to calibrate the models, the proposed algorithm can be trained over different ranges of intervals prior to incidents and can then be deployed to monitor different time periods of the day to detect traffic anomalies. We also assess the potential of deploying the algorithm in V2V and V2I environments, where only a fraction of individual vehicle information might be available at a given time due to limitations in the transmission mediums and/or loss of data due to transmission delays [10].

This paper is organized as follows. A review of related studies is provided in Section II. Section III describes the assessment framework, and the proposed detection methodology is presented in Section IV. In Section V, we use real-world data to demonstrate the effectiveness of our proposed algorithm in the detection and classification of anomalies, and final remarks and future work are outlined in Section VI.

II. Related Work

A. Anomaly Detection and Classification Using Macroscopic Traffic Variables

Macroscopic traffic variables that represent aggregated behavior of vehicles are derived using an analogy with fluid dynamics, and the relationships between these variables are described in the classical Lighthill–Whitham–Richards model [11]. Basic macroscopic traffic variables, derived from inductive loop detectors, notably flow and occupancy, have extensively been employed for traffic incident detection, focusing on detection after a major disruption of traffic has occurred [7], [12]. Recently, a number of studies have suggested that it is worth attempting to detect traffic anomalies prior to the occurrence of an incident [4], [6]. The majority of these studies have shown that measurements of speed deviation can be used as precursor signals of a traffic incident. Unfortunately, one of the main problems with all these approaches is that measuring the variation of speed at a specific location lacks the microscopic-level characteristics that could capture individual vehicle interactions over time; the spatial–temporal microscopic characteristics will be lost once the vehicle passes the detector location.

B. Anomaly Detection and Classification Using Microscopic Traffic Variables

Microscopic traffic variables describe individual vehicle behaviors and their interactions and are known to provide fine-grained information about individual vehicles ([11], [13]). However, few studies have employed microscopic traffic variables for anomaly detection. The study in [14] is among the first to use relative speed and intervehicle spacing to derive a reliability model for freeway traffic flow, but the model itself is used with macroscopic traffic variables, e.g., flow and density, and no algorithm is suggested for anomaly detection.

Relative speed, intervehicle spacing, intervehicle time gap, and lane-change tracking are microscopic traffic variables that have been used for anomaly detection [15]–[17]. In [15], lane-changing fractions that were estimated from loop detectors are used for incident detection, but the effectiveness of this approach depends on the loop detector locations. In VGrid [16], each vehicle uses only its local information to determine whether it is in a queue and, if so, communicate this information to other vehicles. A more recent system called WILLWARN uses onboard sensors to measure microscopic information (e.g., wheel speed and reduced friction) to detect possible hazards [18]. However, the information shared among vehicles is normally restricted to hazard-warning messages. Both VGrid and WILLWARN are not explicitly assessed in terms of FAR as to whether they can effectively use shared microscopic traffic variables for anomaly detection and classification.

Recently proposed anomaly detection systems, i.e., VII-SVM, VII-ANN [19], and NOTICE [17] have been designed to use the speed profile and lane-changing behavior of individual vehicles. However, to obtain such fine-grained information, these systems require a specific infrastructure that consists of sensors and wireless transceivers that are uniformly installed on each road segment [17], [19] and/or on each lane [17]. Such requirements are likely to limit the deployment of VII-SVM, VII-ANN, and NOTICE, because they are not scalable with typical traffic-monitoring systems.

We recently proposed two algorithms that use the variability of microscopic traffic variables from V2V and V2I communications for anomaly detection, even when the information is available only from a fraction of vehicles on the segment [1], [9]. In [9], the algorithm uses relative speed and lane-changing trajectories to detect anomalies that were caused by blocked lanes, and this algorithm outperforms the algorithm in [15] under low vehicle density. We have also shown that the algorithm in [1] can detect anomalies in a real-world data set and outperforms the algorithms in [4] and [15], but its performance misses certain anomalies that occur under high vehicle density.

III. Analysis Framework

In this paper, we introduce a distributed traffic-monitoring framework that can be supported by the information shared among neighboring vehicles. This information would enable the calculation of microscopic traffic variables statistics to detect and classify anomalies on a short road segment (e.g.,

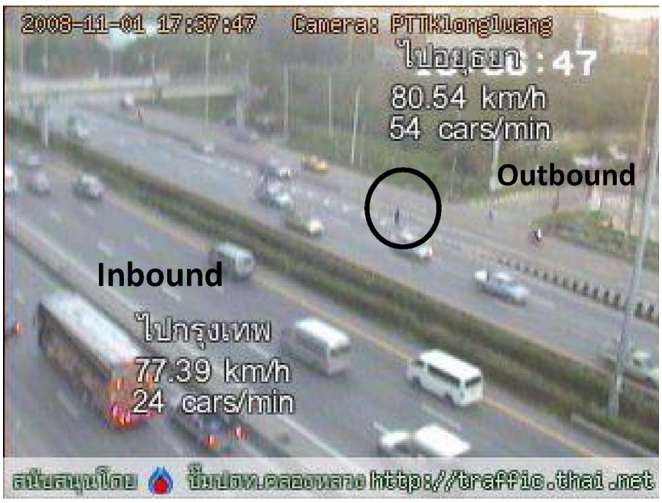

Fig. 1. Transient anomaly caused by a pedestrian on the shoulder of a freeway segment.

Fig. 2. Disruption precursor caused by a truck that blocked a lane.

100–200 m [20]) and within a short time interval (e.g., 5–30 m [21]). The relevant information (e.g., instantaneous speed and position) would be measured by vehicles that share information through automotive navigation systems and wireless communications [16], [22]. Alternatively, the microscopic traffic information could be inferred from currently available roadside infrastructure, e.g., video closed-circuit television surveillance cameras [23].

In the proposed framework, an anomaly detection model $H = \{F, C, D\}$ is stored and can be locally activated when vehicles are present on the segment of interest. H generally consists of the following three entities, which have been trained with historical data: 1) F denotes a feature extraction model; 2) C denotes a classification model; and 3) D denotes a decision-making model. The parameters of model H can be downloaded by the vehicles on the segment or used by a local roadside infrastructure itself for anomaly detection. The input to the model consists of the microscopic traffic variables that can be calculated and shared among individual vehicle and the closest vehicle(s) downstream whose information is available at the time of interest.

In this paper, we consider the following two types of anomalies: 1) *transient anomalies*, and 2) *disruption precursors.* Transient anomalies are defined as deviations of traffic patterns that may be followed by minor disruptions to the traffic flow. One example of a transient anomaly is shown in Fig. 1, where an unexpected appearance of a pedestrian on the freeway shoulder causes distraction and a temporary drop in speed. Disruption precursors are defined as traffic patterns that may lead to a major disruption of the underlying traffic flow. This type of anomaly has received much attention, primarily in cases where it is associated with accidents [4], crashes [6], or congestions [5], [7]. Fig. 2 shows an incident that has evolved from a lane blocking and disrupts the traffic flow. The aim in this paper is to develop an algorithm that can identify the onset of both types of traffic anomalies as the first step toward detecting, classifying, and predicting the impact of the incident so that an appropriate, proactive, neutralizing, and coordinated set of actions can be initiated.

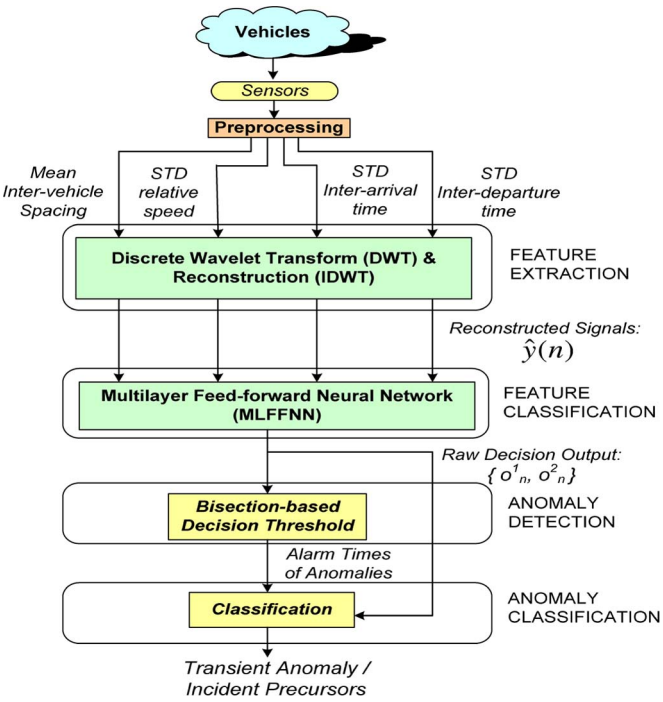

Fig. 3. Proposed anomaly detection and classification methodology.

IV. Methodology

A. Overview of the Proposed Algorithm

The proposed algorithm is based on the observation that a traffic anomaly can be detected by monitoring the changes in the behavior of individual vehicles (e.g., deceleration and lane change), which would be reflected in changes in the variability of the observed microscopic traffic variables [1]. More recently, it has also been suggested that microscopic traffic variables analysis could be used to identify incident precursor phases [8].

The proposed anomaly detection scheme can be divided into major blocks or stages, as depicted in Fig. 3. The process starts with the feature extraction stage, which performs the conversion of the original traffic variables into features that are abstract representation of those variables and contain all the essential

information for the detection task. In this paper, the feature extraction step is based on the use of a wavelet-based filter. For that purpose, each variable is initially decomposed into low- and high-frequency components. Then, the small coefficients of the high-frequency components that are less correlated with the original microscopic signals are removed, and finally, the filtered signal is reconstructed. In the next block, referred to as *feature classification*, every new measurement of the microscopic variables is assigned to a specific class according to the characteristics of the reconstructed filtered signals. In this paper, a multilayer feedforward neural network (MLFFNN) is selected for the feature classification model based on the previous performance [24], [25]. The reconstructed time series of microscopic statistics are then used as inputs for the MLFFNN. The outputs of MLFFNN represent degrees of likelihood between 0 and 1, and hence, a threshold is needed to discover and classify an emerging anomaly pattern. Here, we use a bisection-based decision threshold [26], which is optimized over different traffic conditions to enhance the adaptability of the proposed algorithm. An alarm is raised if the MLFFNN outputs exceed the bisection-based decision threshold for a certain number of consecutive samples [persistency (PT)]. Finally, the alarm times and the outputs from MLFFNN are used to classify the detected traffic anomalies.

B. Inputs Description

The microscopic traffic variables used in this paper for anomaly detection are listed as follows.

- *Intervehicle spacing*: $s_{i,n} = x_{i-1,n} - x_{i,n}$, where $x_{i,n}$ is the position of a vehicle i at time n;
- *Relative speed*: $v_{i,n} = w_{i-1,n} - w_{i,n}$, where $w_{i,n}$ is the speed of a vehicle i at time n;
- *Interarrival time* a_i: The interarrival time is defined as the difference between the arrival time to the *beginning* of a segment of interest of a vehicle i and of the previous vehicle $i-1$ that has arrived;
- *Interdeparture time* d_i: The interdeparture time is defined as the difference between the arrival time to the *end* of a segment of interest of a vehicle i and the arrival time of the previous vehicle $i-1$.

We select as statistics of interest the standard deviations of relative speed, interarrival time, and interdeparture time of vehicles to track changes in variability. To further enhance the learning capabilities of the algorithm under different traffic conditions and with different vehicle density, we select the mean of intervehicle spacing to track changes in vehicle density [27].

C. Feature Extraction

The effectiveness of the anomaly detection algorithm is first determined by the representativeness of the features and how efficiently they can be extracted and identified. Our analysis with a real-world data set in [1] has shown that some anomalies may have weak amplitudes, and in some cases, the boundary between normal and anomalous behavior cannot clearly be distinguished. In addition, changes that are associated with anomalies can occur under different time scales, depending on the underlying traffic conditions. Because we intend to deploy the algorithm in nodes that might have limited storage and communication capabilities, the feature extraction model has to be minimal. Hence, the process of feature extraction is performed by the use of wavelet transforms, which is a well-known technique for the analysis of signals with multiscale behavior [2]. Discrete wavelet transform (DWT) is chosen, because it can represent information signals with orthogonal wavelet basis, which minimize the number of associated representative coefficients [28]. Furthermore, DWT can represent the original signal using wavelet detail coefficients (fine-scale information) and approximation coefficients (coarse-scale information), which would enhance the ability to extract changes in microscopic traffic variables associated with different time scales. In the proposed scheme, DWT is used to individually filter relevant components of the four input microscopic traffic variables described in Section IV-B. The output of the filtering process is the reconstruction of each of these microscopic traffic variables using only the approximation coefficients and the wavelet coefficients that are highly correlated with the original signal.

Based on the DWT framework, a signal $y(n)$ is decomposed into levels of approximations and details to form a multiresolution analysis of the signal as follows: $y(n) = \sum_k a_{J,k}\phi_{J,k}(n) + \sum_{m\geq J}\sum_k d_{m,k}\psi_{m,k}(n)$, where J, m, $k \in \mathbb{Z}$, $a_{J,k}$ denotes the approximation coefficient at resolution J, $d_{m,k}$ denotes the wavelet coefficient at resolution m, $\phi_{J,k}(n)$ is a scaling function, and $\psi_{m,k}(n)$ is a wavelet function at resolution m. The coefficients $a_{L,k}$ and $d_{m,k}$ are computed as follows: $a_{J,k} = \langle y(n), \phi_{J,k}(n)\rangle$, $d_{m,k} = \langle y(n), \psi_{m,k}(n)\rangle m$, $k \in \mathbb{Z}$, where the operator $\langle . \rangle$ denotes the inner product in the space of square integrable functions $J^2(\mathbb{R})$. The dyadic DWT assumes scaling functions ϕ and wavelet functions ψ of the following form: $\phi_{m,k}(n) = 2^{m/2}\phi(2^m n - k)$, $\psi_{m,k}(n) = 2^{m/2}\psi(2^m n - k)m$, $k \in \mathbb{Z}$. In particular, DWT is used to obtain the approximation coefficients $a_{L,k}$ and wavelet coefficients $d_{m,k}$, whereas the process of reconstructing the signal, given such coefficients, is called the inverse DWT. In practice, the dyadic DWT can be implemented in a computationally efficient manner using the dyadic filter tree algorithm [28].

D. Feature Classification

Identifying anomalies can be a difficult task due to possible nonlinear relationships between patterns of microscopic traffic variables and classes (normal, transient anomaly, and disruption precursors) [1]. These relationships could become even more complex when there is a change in the underlying traffic condition, because a sample of microscopic traffic variables that are associated with one class in one traffic condition may belong to another class in another traffic condition. To capture such nonlinear relationships, an MLFFNN is selected for feature classification, because once properly trained, MLFFNN has often been found to perform well in classifying traffic patterns when deployed on site [24], [25], [29].

For a classification problem, the primary task of MLFFNN is to learn a classification function as follows: $C : \underline{Y} \rightarrow \underline{O}$, where

$\underline{Y} = \{y_n^i\}$ is a set of input to the ith node in the input layer of MLFFNN, and $\underline{O} = \{o_n^k\}$ is an output vector from the kth node in the output layer at time n [30]. Let w^{ij} denote the weight for the connection from node i in the input layer to node j in the hidden layer and v^{jk} denote the weight for the connection from node j in the hidden layer to node k in the output layer. A sigmoid function, $SIG(x) = 1/(1+e^{-x})$, is used as a transfer function in the nodes in both the hidden and output layers, which can be represented as

$$z_n^j = SIG\left(\sum_i w^{ij} y_n^i\right) \tag{1}$$

$$o_n^k = SIG\left(\sum_j v^{jk} z_n^j\right). \tag{2}$$

Based on the preliminary assessment of the algorithm with the training data set, MLFFNN with two hidden layers is chosen for feature classification. Let m_0, m_1, m_2, and m_3 denote the number of neurons in the input layer, the first hidden layer, the second hidden layer, and the output layer of the proposed algorithm, respectively. Because there are four microscopic traffic variables for input, the input layer consists of four neurons, i.e., $m_0 = 4$. The number of neurons in each hidden layer is chosen as a function of the number of microscopic traffic variable inputs, with the aim of balancing between not having too few neurons (e.g., the capability to model nonlinear mapping) and not having too many neurons (e.g., the problems of excessive time-consuming and having too many local minimums). We choose the same set up as shown in [25] and [31], where $m_1 = m_2 = 3 * m_0 + 1$. Each element in the output of the MLFFNN, $\{o_n^1, o_n^2\}$, is set to be between 0 and 1, which would require a decision-making threshold to classify whether the output corresponds to an anomaly.

E. Decision Algorithm for Anomaly Detection

It is crucial that, during the training process, the proposed algorithm can examine a range of possible output values to find an optimal decision-making threshold for assessing the outputs of MLFFNN. We employ a threshold-varying bisection method in [26], which can be summarized as follows. During the training process, a range of thresholds $[\alpha, \beta]$, $0 \le \alpha \le \beta \le 1$ is considered, where the goal is to find a threshold γ that minimizes the cost function f_γ. At each step, f_α, f_β, and $f_{(\alpha+\beta)/2}$ are compared. If $f_{(\alpha+\beta)/2} < \min(f_\alpha, f_\beta)$, then the algorithm selects $\gamma = (\alpha+\beta)/2$ and exits; otherwise, it repeats the previous step, but with a bisected threshold range $[(\alpha+\beta/2), \beta]$ if $f_\alpha > f_\beta$ or $[\alpha, (\alpha+\beta/2)]$ if $f_\alpha \le f_\beta$. The bisection process is repeated until the minimum cost function is reached, i.e., $f_\gamma < \min(f_\alpha, f_\beta)$, or the current lower bound is at least ϵ less than the current upper bound, i.e. $\beta - \alpha \le \epsilon$. In the latter case, the algorithm selects $\gamma = \beta - \epsilon$ as its final threshold, which is used to assess the raw decision output from MLFFNN. In the proposed scheme, a lower bound of $\alpha = 0.3$ and an upper bound of $\beta = 0.7$ are chosen, because a larger initial interval (e.g., [0.1, 0.9]) could cause the bisection method to select the threshold γ that is too biased toward a training set dominated by a particular traffic condition (e.g., too close to 0.1 or 0.9) [26].

To enable the decision algorithm to select a threshold γ that maximizes the DR and minimizes the FAR, we use the cost function $f_\gamma = w \times FAR_\gamma + (1-w) \times (1 - DR_\gamma)$ for the modified bisection, where FAR_γ = FAR, and DR_γ = DR when a threshold γ is used. Hence, the term $1 - DR_\gamma$ corresponds to the number of missed detections. The choice of w $(0 \le w \le 1)$ depends on whether maximizing the DR or minimizing the FAR is more important, which subsequently depends on practical considerations and the purpose of the observations. In this experiment, the value of 0.5 is chosen for w to balance between maximizing the DR and minimizing the FAR.

The raw decision output from MLFFNN, as shown in Fig. 3, consists of two separate signals, $\{o_n^1, o_n^2\}$, whose binary combination, at time n, determines whether an anomaly has occurred, as well as the type of the detected anomaly: {0, 0} for normal traffic patterns, {0, 1} for transient anomalies, and {1, 1} for disruption precursors. For the anomaly detection decision, PT is used to determine the number of consecutive samples of the raw decision output, $\{o_n^1, o_n^2\}$, that exceed the threshold. The alarm is raised if at least one element of $\{o_n^1, o_n^2\}$ consecutively exceeds the threshold for a number of PT samples.

F. Decision Algorithm for Anomaly Classification

The classification decision function is based on the assessment of the likelihood that an anomaly corresponds to a transient anomaly or a disruption precursor. In the proposed algorithm, the assessment of likelihood is performed on each raw decision output of the MLFFNN, which is modeled as a classifier whose value represents likelihood [32]. The likelihood combination techniques in [32] are originally proposed to combine outputs over space from different NNs. In this paper, we apply the same concept to combine the raw decision output over time from the same NN, i.e. for a given time n, the raw decision output $\{o_1, o_2\}_n$ is itself a classifier. The classification decision function $f(\eta)$ can be formalized as shown in (3), where the raw decision outputs of MLFFNN are monitored from the time of detection n_0 to $n_0 + \eta$, where η denotes a classification interval, i.e., $\{o_1, o_2\}_{n_0}, \{o_1, o_2\}_{n_0+1}, \ldots, \{o_1, o_2\}_{n_0+\eta}$

$$f(\eta) = \arg\max_c \sum_{n=n_0}^{n_0+\eta} w_n^c f_n^c, \quad c = 1, 2. \tag{3}$$

The weight w_n^c is the weighted average of the likelihood f_n^c, i.e., $w_n^c = f_n^c / (\sum_{c=1,2} f_n^c)$. The likelihood f_n^c is calculated from the raw decision outputs $\{o_n^1, o_n^2\}$. Recall that the raw decision output o_n^1 nearer to zero indicates more likelihood of transient anomaly, whereas the raw output o_n^1 closer to one indicates more likelihood of disruption precursors. Hence, f_n^c is calculated as follows: $f_n^1 = (1 - o_n^1) \times o_n^2$, and $f_n^2 = o_n^1 \times o_n^2$.

G. Distributed Deployment of the Proposed Algorithm

The proposed algorithm is designed such that the anomaly detection and classification functions are performed in a distributed manner by neighboring vehicles and/or by a local

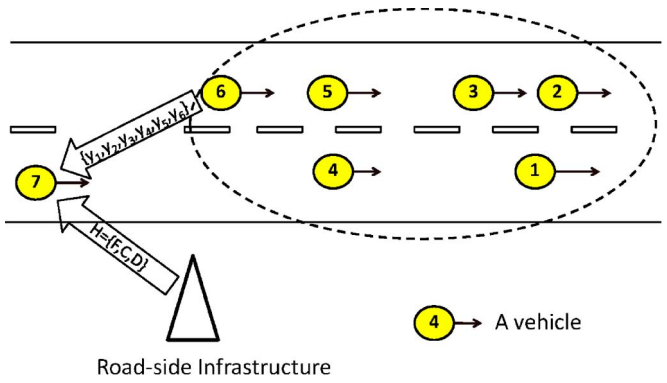

Fig. 4. Example of distributed deployment of the proposed algorithm.

roadside infrastructure based on the locally collected information. Fig. 4 shows an example of how microscopic traffic information of individual vehicles can be used by the proposed algorithm to detect traffic anomalies on the road segment of interest. In this example, the proposed algorithm is used by vehicle 7 to identify traffic anomalies based on microscopic traffic information from downstream vehicles 1–6. In this setting, each vehicle learns microscopic traffic variables of other vehicles by exchanging information about itself with other vehicles in the neighborhood. In the figure, vehicle 6 collects and sends to vehicle 7 microscopic traffic variables $\{y_i, i = 1, \ldots, 6\}$ (speeds, positions, arrival, and departure times) of other vehicles on the segment, where the mean of intervehicle spacing and the standard deviations of relative speed, interarrival time, and interdeparture time are calculated based on the received microscopic information. The anomaly detection and classification model H for the road segment is downloaded by vehicle 7 from, for example, a roadside infrastructure.

V. Performance Evaluation Using Real-World Data

A. Performance Evaluation Parameters

Let $n_{p,q}$ be the time that the pth alarm is raised for an anomaly recorded at a time point n_q. The alarm is considered true if $n_{q,p} \in [n_q, n_q + n_b]$, ; otherwise, it is considered a false alarm, in which case n_b denotes a detection interval used for evaluation purposes only and should not exceed anomaly duration. Given M anomaly cases in the experiment, the performance evaluation parameters that we consider are the DR, MTTD, and FAR, i.e.,

$$\text{DR} = \frac{\text{Number of anomalies detected}}{M} \tag{4}$$

$$\text{MTTD} = \frac{\sum_{p=1}^{M}(n_{q,p} - n_q)}{M}, \quad n_{q,p} \in [n_q, n_q + n_b] \tag{5}$$

$$\text{FAR} = \frac{\text{Number of alarms not in } [n_q, n_q + n_b]}{\text{Total number of alarms}}. \tag{6}$$

B. Descriptions of the Data

To assess the algorithm, we chose a real-world freeway segment in which the microscopic traffic variables can also be obtained from a video surveillance camera (see Figs. 1 and 2). The road segment that we studied is part of a freeway network that links Bangkok to the northern provinces of Thailand. During the five-month period of data collection, the free-flow maximum and average speeds on this freeway segment were found to be 120 and 80 km/h, respectively. The nearest junctions are approximately 1.2 km upstream and 1.5 km downstream. Recurring traffic congestion occurs during morning rush hours, between 7 and 9 A.M., where congestion usually originates at the downstream junction and propagates upstream to the freeway segment studied during these periods. We note that these recurring congestions are not included as anomaly cases in this paper.

1) Calculation of Traffic Variables: The image processing software used on this camera to track individual vehicles has been developed by a team of researchers and is reported in [23]. The object-tracking accuracy of this camera has been estimated to be approximately 80%. Although the proposed algorithm was initially assessed using data from video camera, it can also be deployed with the support of a vehicular ad hoc network (VANET)-based information-sharing system in which individual vehicle tracking is performed by automatic vehicle location systems [e.g., Global Positioning System (GPS)], real-time location systems (e.g., radio frequency ID), or inertial navigation systems (e.g., dead reckoning), and the information can be shared using wireless communication technologies such as cellular GPRS, radio wave, or infrared [16], [22], [33]. A recent study in a simulated environment reported in [34] has shown that vehicular speed can be estimated using onboard vehicle data, with error as low as 10.5%.

On the image frame of the camera, as shown, for example, in Fig. 2, virtual entrance and exit lines were drawn at the beginning and end of the segment, respectively. For a vehicle i, $\{t_i^{in}, t_i^{out}, w_i^{\text{emp}}\}$ was recorded, where t_i^{in} is the time that the vehicle crossed the entrance line, t^{out} is the time that the vehicle crossed the exit line, and w_i^{emp} is the speed of vehicle i, calculated as $w_i^{emp} = \text{Segment Length}/(t_i^{out} - t_i^{in})$. We note that the accuracy of individual vehicle speed will not have a significant impact on the performance of the proposed algorithm, because it does not rely on measurements of individual vehicles to detect anomalies (see Section IV-B).

The interarrival time is calculated as $a_i^{emp} = t_i^{in} - t_{i-1}^{in}$, whereas the interdeparture time is calculated as $d_i^{emp} = t_i^{out} - t_{i-1}^{out}$. For a sampling interval of τ (in seconds), the averages of the interarrival time $(\overline{a_{m\tau}^{emp}})$ and the interdeparture time $(\overline{d_{m\tau}^{emp}})$ and the standard deviations of the interarrival time $(\sigma_{a,m\tau}^{emp})$ and the interdeparture time $(\sigma_{d,m\tau}^{emp})$ of the sampling interval $m\tau$ can be calculated as follows: $\overline{a_{m\tau}^{emp}} = (\sum_{i=1}^{I} a_i^{emp})/I$, $\sigma_{a,m\tau}^{emp} = \sqrt{(\sum_{i=1}^{I}((a_i^{emp})^2 - (\overline{a_{m\tau}^{emp}})^2)^2)/I}$ for $(m-1)\tau \le t_i^{in} < m\tau$, and $\overline{d_{m\tau}^{emp}} = (\sum_{i=1}^{I} d_i^{emp})/I$, $\sigma_{d,m\tau}^{emp} = \sqrt{(\sum_{i=1}^{I}((d_i^{emp})^2 - (\overline{d_{m\tau}^{emp}})^2)^2)/I}$ for $(m-1)\tau \le t_i^{out} < m\tau$, where $m = 1, 2, 3, \ldots$.

Because the camera cannot yet directly measure spaces between individual vehicles, the average intervehicle spacing is instead calculated from the average interdeparture time and average speed as follows: $\overline{s_{m\tau}^{emp}} = \overline{d_{m\tau}^{emp}} \times \overline{w_{m\tau}^{emp}}$, where

$\overline{w_{m\tau}^{emp}} = (\sum_{i=1}^{I} w_i^{emp})/I$. A relative speed observed by vehicle i to its leading vehicle $i-1$ is calculated as $v_i^{emp} = w_{i-1}^{emp} - w_i^{emp}$ for $t_i^{out} > t_{i-1}^{out}$. Then, for a sampling interval of τ (in seconds), the average $(\overline{v_{m\tau}^{emp}})$ and standard deviation $(\sigma_{m\tau}^{emp})$ of relative speeds of the sampling interval $m\tau$ can be calculated as follows: $\overline{v_{m\tau}^{emp}} = (\sum_{i=1}^{I} v_i^{emp})/I$, and $\sigma_{m\tau}^{emp} = \sqrt{(\sum_{i=1}^{I}((v_i^{emp})^2 - (\overline{v_{m\tau}^{emp}})^2)^2)/I}$ for $(m-1)\tau \leq t_i^{out} < m\tau$, where $m = 1, 2, 3, \ldots$.

2) Obtaining Anomaly Cases: The real-world data set used in this paper was collected by the camera under different traffic conditions daily from 6 A.M. to 6 P.M. over a five-month period from August to December 2008, plus an additional two months of video postprocessing to visually identify anomaly cases. The period of five months obtained sufficient anomaly cases for performance evaluation. The average number of vehicles detected that pass the freeway segment is approximately 30 000 vehicles per day; therefore, there are approximately 5 000 000 records for the five-month period of data collection. For evaluation purposes, anomalies that took place were independently logged by a team of transportation researchers using video images from the camera at the target site. The video images were used to log anomalies that were identified as disruption precursors when they were associated with major traffic disruption; otherwise, the anomalies were recorded as transient anomalies. There are 26 anomaly cases in the data; 9 cases were transient anomalies, and 17 cases were disruption precursors. These anomalies took place under various traffic conditions such as low-flow high-speed, high-flow high-speed, and low-flow low-speed conditions.

Based on video images, transportation researchers were also asked to log the start and end times of the anomaly cases. Each anomaly case consists of a set of timestamps $\{T_s^a, T_s^i, T_e^i\}$, where T_s^a denotes the time when a traffic anomaly is recorded (e.g., a pedestrian and a lane-blocking truck are observed in Figs. 1 and 2, respectively), T_s^i denotes the times when a traffic incident is recorded to take place, and T_e^i denotes the end time of traffic incident or when traffic has recovered. For transient anomalies, only T_s^a and T_e^i are recorded, because incident occurrence is not observed. The duration of traffic anomalies analyzed in this paper ranges from 5 min to 45 min. In the performance evaluations in Section V-D and V-E, an anomaly is considered detected if an alarm is raised by the algorithm within $[T_s^a, T_s^i]$ for disruption precursors and $[T_s^a, T_e^i]$ for transient anomalies. Then, the MTTD is calculated as the average of the difference between the alarm time and T_s^a.

C. Benchmark Anomaly Detection and Classification Algorithm

We chose the algorithm proposed in [1] as our benchmark, because it has been shown to outperform the algorithms in [4] and [15]. The benchmark algorithm combines decision with regard to temporal changes in the variances of microscopic variables such as relative speed and interarrival and interdeparture of vehicles, with spatial changes related to upstream and downstream traffic patterns. An alarm was activated using a weighted-vote scheme.

TABLE I
PERFORMANCE COMPARISON RESULTS OBTAINED BY AVERAGING 20 EXPERIMENTS FOR DIFFERENT VALUES OF PT

Proposed Algorithm (Figure 3)	DR	FAR	$MTTD$
PT = 1	1.00	0.057	469.9
PT = 2	1.00	0.050	516.3
PT = 3	0.95	0.035	516.4
Benchmark Algorithm [1]	DR	FAR	$MTTD$
PT = 1	0.88	0.370	370.4
PT = 2	0.73	0.170	139.1
PT = 3	0.69	0.150	373.1

D. Anomaly Detection Assessment

The experiment reported in this section is conducted using a cross-validation technique. Anomaly cases in our data were separated into disjoint training and testing sets; the training set consists of 12 anomaly cases, and the testing set consists of 14 anomaly cases. The training set was used to train the MLFFNN and find the optimal threshold by the bisection method in the proposed algorithm, whereas the testing set was used for performance evaluations. In each experiment, the anomaly cases in the training and testing sets were randomly selected from the 26 anomaly cases in our data set. Furthermore, to reflect real-world operational conditions, each algorithm was continuously assessed from 6 A.M. to 6 P.M. and for each day of the data set.

Table I shows performance evaluation results for the DR, FAR, and the MTTD, obtained by averaging over 20 experiments. The detection interval n_b of each anomaly case in (4)–(6) is the recorded duration of that anomaly case, which is found to be between 5 and 45 min.

The proposed algorithm detects a relatively higher number of anomalies with a much smaller FAR than the benchmark algorithm. Furthermore, the proposed algorithm detects anomalies that are missed in [1]. We note that the missed anomaly cases took place while there was already a high flow of vehicles ($\geq$ 2000 vehicles/h) on the segment and where the boundary between the variability of relative speed associated with normal and anomalous traffic conditions cannot clearly be distinguished.

The proposed algorithm successfully classifies anomaly cases that cannot be detected by the benchmark algorithm [1], and this is reflected in the slightly higher MTTD values shown in Table I. These anomalies are associated with a high volume of vehicles on the freeway segment, where the variability of the microscopic variables is bounded by the lack of room to maneuver vehicles. Consequently, it takes longer for the proposed algorithm to detect subtle changes in microscopic variability. Furthermore, we note that MTTDs shown in Table I are the average detection delays of traffic anomalies and not the average detection delays of traffic incidents. All the anomalies included in the test data set are detected before the actual (independently recorded) traffic incidents time are logged and/or before the observed traffic patterns settled back to, for example, normal free-flow conditions. Further analysis on these real-world data sets shows that the alarm times of the proposed algorithm are, on the average, triggered 13 min before the occurrence of incidents, which should still give enough time to initiate an appropriate response.

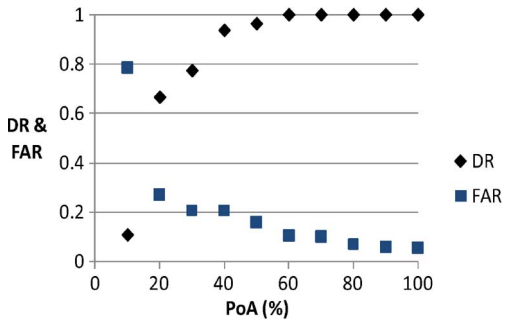

Fig. 5. DRs and FARs under different PoAs.

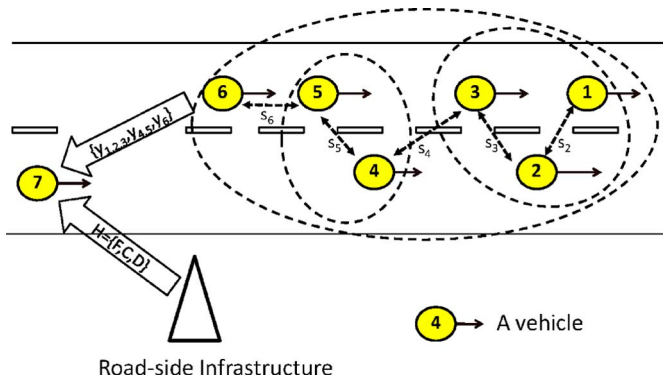

Fig. 6. Example of distributed deployment of the proposed algorithm under information aggregation using, for example, the method in [10].

E. Anomaly Detection With Partial Coverage of Vehicle Information

1) Anomaly Detection With Partial Availability of Data From Vehicle Population: The statistics of microscopic traffic variables are calculated from a percentage of the vehicles population $I = PoA * I_{total}$, where PoA is the percentage of availability, and I_{total} is the total number of vehicles on the segment. Fig. 5 shows that the DR and FAR when the proposed algorithm has access to PoA range from 10% to 100%. The proposed algorithm can detect all anomaly cases with the FAR less than 20% for 50% $\leq$ PoA $\leq$ 100%. Because the average vehicle density I_{total} was observed to be between 6 and 20 vehicles/lane, the results in Fig. 5 shows that the proposed algorithm can still identify anomalies in the real-world data set, even when the microscopic traffic information is available only for 50% of vehicles on the segment, (i.e., for I as low as three vehicles per lane). As PoA is further reduced to the range $\leq$ 40% (i.e., $I \approx 2-3$ vehicles/lane), the information is no longer sufficient to identify all traffic anomalies. Hence, to keep the DR and FAR at acceptable levels, the proposed algorithm has to particularly be trained for low PoA with a sufficient number of anomaly cases.

2) Anomaly Detection With Aggregated Information: In this experiment, a ratio-based aggregation algorithm adopted from the TrafficView system [10] is employed to form clusters of vehicles where microscopic traffic variables of individual vehicles are locally aggregated according to the availability of medium access control (MAC) payload size, the amount of microscopic traffic information, and the relative distances of vehicles with respect to one another.

In the ratio-based algorithm, a cluster is formed, where every vehicle makes itself known to other vehicles, and the upstream-most vehicle in the cluster is selected as a cluster head. The cluster head determines if the remaining MAC payload size R is enough for dissemination of microscopic traffic variables individually. If the size R is not enough, then local aggregations are needed, and subclusters are formed according to the relative distance between each vehicle to the upstream cluster head. An example of how microscopic traffic variables are aggregated when the available MAC payload size is reduced to $0.3R$ is shown in Fig. 6. Let us hypothetically assume that $0.1R$ is needed to transmit a microscopic traffic variable of each vehicle so that the available bandwidth can accommodate only a single vehicle information from each subcluster. In this example, vehicle 6 acts as a cluster head, which collects the aggregated information from the two downstream subclusters, vehicles {1, 2, 3} and vehicles {4, 5}. Let y_i denote a microscopic traffic variable measured by vehicle i and s_i denote intervehicle spacing between vehicle i and the closet vehicle downstream. The aggregated microscopic traffic information sent to vehicle 7 would be the weighted average $\{y_{1,2,3}, y_{4,5}, y_6\}$, where $y_{1,2,3} = (\sum_{i=1}^{3} s_{i,6}y_i)/(\sum_{i=1}^{3} s_{i,6})$, $y_{4,5} = (\sum_{i=4}^{5} s_{i,6}y_i)/\sum_{i=4}^{5} s_{i,6})$, and $s_{i,6} = \sum_{j=1}^{6} s_j$. Then, vehicle 7 would use $\{y_{1,2,3}, y_{4,5}, y_6\}$ to calculate the microscopic statistics.

TABLE II
ASSESSMENT OF THE PROPOSED ALGORITHM UNDER DIFFERENT PoA AND PERCENTAGES OF MAC PAYLOAD AVAILABILITY (R) IN THE REAL-WORLD DATA SET

PoA (%)	Unlimited R		$0.5R$		$0.3R$	
	DR	FAR	DR	FAR	DR	FAR
100	1.000	0.057	1.000	0.058	1.000	0.059
90	1.000	0.058	1.000	0.059	1.000	0.064
80	1.000	0.061	1.000	0.070	1.000	0.079
70	1.000	0.076	1.000	0.101	1.000	0.107
60	1.000	0.104	1.000	0.104	1.000	0.161
50	1.000	0.161	0.955	0.173	0.955	0.187
40	0.937	0.197	0.937	0.206	0.921	0.207
30	0.767	0.197	0.774	0.206	0.671	0.228
20	0.666	0.282	0.648	0.289	0.572	0.307
10	0.108	0.504	0.109	0.788	0.060	0.870

Table II shows the DR and FAR when the proposed algorithm is applied to detect anomalies in the real-world data set.[1] The MTTD is not shown in this table, because it is found to have consistent values of approximately 8–9 min for every combination of PoA and R analyzed. The DR and FAR are averaged over ten experiments, where in each experiment, the anomaly cases in the training and testing sets are randomly selected from the 26 anomaly cases in our data set.

Table II shows that the proposed algorithm detects anomalies with more than 90% DR and FAR less than 20%, even when PoA is as low as 40% and only 50% and 30% of MAC payload ($0.5R$ and $0.3R$) are available (information obtained from only two to thre vehicles). It is also shown that the reduction of the MAC payload size does not have a significant impact on either the DR or the FAR. However, when $PoA = 10\%$, as previously

[1] The MAC payload size R is set to be 2313 B for an 802.11b-based wireless network, and the average size for a single vehicle record is 50 B, as shown in [10].

TABLE III
ASSESSMENT OF THE PROPOSED ALGORITHM ON THE REAL-WORLD DATA SET, WITH PACKET TRANSMISSION DELAYS, $PoA = 50\%$, AND MAC PAYLOAD SIZE $= 0.5R$. THE TIME WINDOW SIZE IS 300 S. EXP REFERS TO THE EXPONENTIAL DISTRIBUTION, AND GP REFERS TO THE GENERALIZED PARETO DISTRIBUTION

Packet Transmission Delay (s)	DR	FAR	MTTD(s)
No Delay	0.955	0.173	448.2
EXP with mean 202.9s	0.929	0.199	563.6
EXP with mean 302.9s	0.854	0.257	616.8
GP with mean 202.9s	0.909	0.141	515.3
GP with mean 302.9s	0.854	0.230	620.0

TABLE IV
CLASSIFICATIONS ON REAL-WORLD DATA FROM 200 EXPERIMENTS, CC = CORRECTLY CLASSIFIED, MC = MISCLASSIFIED, $\eta = 5$ min.

Proposed Algorithm Unlimited R, $PoA = 100\%$	Number of Cases	CC Cases	MC Cases
Transient Anomalies	9	9	0
Disruption Precursors	17	17	0
Benchmark Algorithm [1]	Number of Cases	CC Cases	MC Cases
Transient Anomalies	9	6	3
Disruption Precursors	14	14	0

highlighted, most anomaly cases cannot be detected, because microscopic traffic variables are obtained from a single vehicle.

3) Anomaly Detection With Packet Transmission Delay: In this experiment, a packet transmission delay is generated from a probability distribution. If the generated packet transmission delay exceeds a time window L, the microscopic traffic information of that vehicle is assumed lost. We use not only an exponential distribution but, as suggested in [35], also include a generalized Pareto distribution in our tests to generate packet transmission delays.

It is shown in Table III that the proposed algorithm performs relatively well, because it detects traffic anomalies with more than 85% DR, even when the mean packet transmission delay (302.9 s) exceeds the time window size. We can see that the FAR increases when the mean packet transmission delays exceed the time window size, because there is not enough microscopic traffic information to distinguish between normal and anomalous traffic patterns. These results again highlight the importance of setting the time window size L to be at least equal to the expected mean packet transmission delay.

F. Anomaly Classification Assessment

Once an anomaly is detected by a vehicle on a particular freeway segment at time n_0, a node (a vehicle or a roadside infrastructure) within the freeway segment would continue to gather more information from the vehicles on the segment and monitor the raw decision outputs $\{o_n^1, o_n^2\}$ from the MLFFNN until $n_0 + \eta$, where η is the classification interval. For the results presented in this section, η is set to 5 m, which is found to give enough classification accuracy in our analysis using the training data set. The classification decision is then made based on the criteria described in Section IV-F.

The classification results in Tables IV and V are derived based on subset of 200 experiments that were selected at random, where in each experiment, the training and testing sets are randomly selected from the 26 anomaly cases in our real-world data set. In Table IV, anomaly cases are considered *correctly classified* (CC) if the proposed algorithm can identify their types within the logged duration, i.e., $[T_s^a, T_s^i]$. The *misclassified cases* (MCs) show the number of anomalies that are incorrectly identified by the proposed algorithm.[2] The proposed algorithm can correctly classify all anomaly cases, whereas the benchmark algorithm [1] misclassifies three transient anomaly cases as disruption precursors. In fact, these cases were identified as distractions on the shoulder of the freeway due to the unexpected appearance of pedestrians in two cases and disabled vehicles in another case (see Fig. 1). These anomaly cases are misclassified by the benchmark algorithm in [1], because the distractions increase the variability of both the relative speed and the covariance between the interarrival and interdeparture times, which subsequently cause the weighted combination of both temporal and spatial alarms to exceed its classification threshold. In contrast, the algorithm proposed in this paper does not rely on any particular threshold, because it compares only the likelihood of the detected anomaly being transient anomaly and the likelihood of being disruption precursor.

TABLE V
CLASSIFICATIONS ON REAL-WORLD DATA FROM 200 EXPERIMENTS WITH DIFFERENT PoA AND PERCENTAGE OF MAC PAYLOAD AVAILABILITY (R), CC = CORRECTLY CLASSIFIED, MC = MISCLASSIFIED, $\eta = 5$ min

Proposed Algorithm	Number of Cases	CC Cases	MC Cases
(Unlimited R, $PoA = 50\%, 100\%$)			
Transient Anomalies	9	9	0
Disruption Precursors	17	17	0
($0.5R$, $PoA = 50\%, 100\%$)			
Transient Anomalies	9	9	0
Disruption Precursors	17	17	0
($0.3R$, $PoA = 50\%, 100\%$)			
Transient Anomalies	9	9	0
Disruption Precursors	17	15	2

We further assess the performance of the proposed algorithm under different PoA coverage and when traffic variables are aggregated according to different percentages of MAC payload availability R. It is shown in Table V that, even when both PoA and MAC payload availability are reduced, all transient anomalies and most of the disruption precursors can still be CC. We note that the MCs are disruption precursors that occurred under moderate vehicle density, where the persistence of variation of microscopic traffic variables causes the proposed algorithm to misclassify these disruption precursors as transient anomalies.

VI. CONCLUSION

This paper proposed a novel anomaly classification algorithm that can be deployed in a distributed manner and utilizes microscopic traffic variables shared by neighboring vehicles to detect and classify traffic anomalies under different traffic conditions. The algorithm uses multiresolution concepts and is based on the likelihood estimation of the outputs of a feedforward neural network and a bisection-based decision threshold.

[2] In [1], we refer to a disruption precursor as an incident precursor. For the benchmark algorithm, the three misdetected cases of incident precursors are excluded from classification evaluation.

The algorithm monitors the change of variability in relative speed, interarrival time, and interdeparture time to capture traffic anomalies. Furthermore, to adapt to underlying traffic conditions, the average intervehicle spacing is monitored.

The algorithm's performance on a real-world data test set presents a significant improvement with respect to the FAR and DR over the benchmark algorithm proposed in [1]. The preliminary results are very encouraging, because they suggest that the algorithm's level of adaptability and resilience to loss of information as a result of low PoA or degradation in communications performance is high. For the data set under analysis, it can also detect anomalies, on the average, 13 min prior to the occurrence of traffic incidents, therefore significantly increasing the time available to initiate appropriate responses.

There are various aspects worth further consideration. One aspect is to design a model that can incorporate information about traffic anomalies from several road links to infer possible impacts on the entire road network. Another aspect is an extension to urban arterial, where changes due to traffic signals and its effect in traffic patterns will need to be considered. It would also be interesting, from a traffic management point of view, to classify traffic anomalies according to physical causes (e.g., lane blocking [9]) to provide more information for incident management. The inclusion of spatial information from other detectors (e.g., cameras on neighboring segments) to locate the bottlenecks and investigate the existence of synchronized flow [36] is also worth investigating. Finally, a more exhaustive assessment of selected VANET technologies and associated protocols with respect to transporting and processing large-scale traffic and sensing information, has yet to be assessed in real-world scenarios and is suggested for further investigation.

APPENDIX
COMPLEXITY ANALYSIS FOR IMPLEMENTATION

The main computational complexity of the proposed algorithm consists of the computations (e.g., additions and multiplications) in the DWT and MLFFNN in Fig. 3, whereas the computational complexity of decision functions for anomaly detection and classification blocks are marginal, because they mainly consist of comparisons. The computational complexity of the DWT depends on the input window size L and the number of coefficients in the wavelet and scaling filters K. The number of additions and multiplications is originally found to be $2K(L+K-1)$ using a shift-register in the well-known Pyramid algorithm, but with a more recently proposed method, it is possible to reduce the number of additions and multiplications to $K*L$ [37].

For the MLFFNN block, it is assumed that the proposed algorithm has been properly trained and the computational complexity is calculated only when the MLFFNN part of the proposed algorithm is downloaded and implemented on a node. The computational complexity of MLFFNN is mainly composed of the number of additions, multiplications, and transfer functions, which subsequently depends on the number of neurons in each layer of the neural network. The number of multiplications, additions, and transfer functions as functions of the number of neurons are $m_0m_1+m_1m_2+m_2m_3$, $m_1+m_2+m_3$, and m_1+m_2, respectively [38] (see Section IV-D for selections of m_0, m_1, m_2, and m_3).

For the benchmark algorithm [1], the computational complexity mainly depends on the eigenvalue decomposition block, whereas the complexity of the Bayesian change detection and the weighted vote scheme blocks are relatively marginal, because they mainly consist of closed-form models and comparisons. The eigenvalue decomposition is commonly obtained by implementing singular value decomposition (SVD). Each small eigenvalue in the input window size L can be calculated by performing SVD on a matrix of size $m \times m$, where m is the number of microscopic traffic variable inputs, and it is well known that the complexity of computing SVD is $O(m^3)$. Therefore, the computation of L small eigenvalues would involve a complexity of approximately $L*O(m^3)$.

Our preliminary analysis shows that, compared to the benchmark algorithm [1], the computational complexity for the implementation of the proposed algorithm is, indeed, higher due mainly to the computations in MLFFNN. However, a number of experiments in the literature have illustrated the feasibility of deploying neural networks in software [21], [39].

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their useful comments on previous versions of this paper and Dr. Wasan and Dr. Supakorn of the National Electronic and Computer Technology Center, National Science and Technology Development Agency, Pathumthani, Thailand, for their assistance in providing the real-world data for the analysis.

REFERENCES

[1] J. A. Barria and S. Thajchayapong, "Detection and classification of traffic anomalies using microscopic traffic variables," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 695–704, Sep. 2011.

[2] V. Alarcon-Aquino and J. A. Barria, "Anomaly detection in communication networks using wavelets," *Proc. Inst. Elect. Eng.—Commun.*, vol. 148, no. 6, pp. 355–362, Dec. 2001.

[3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, Jul. 2009.

[4] C. Oh, J. Oh, and S. G. Ritchie, "Real-time hazardous traffic condition warning system: Framework and evaluation," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 3, pp. 265–272, Sep. 2005.

[5] K. Blake, N. Chaudhary, C. Chu, S. Kuchangi, P. Nelson, P. Songchitruksa, D. Swaroop, and V. Tyagi, "Dynamic traffic flow modeling for incident detection and short-term congestion prediction: 1-year progress report," Texas Transp. Inst., Texas AM Univ., Tamu College Station, TX, Tech. Rep. 0-4946-1, Sep. 2005.

[6] C. Lee, B. Hellinga, and F. Saccomanno, "Real-time crash prediction model for the application to crash prevention in freeway traffic," in *Proc. Transp. Res. Board Nat. Acad.*, Washington, D.C., 2003, pp. 68–77.

[7] E. Parkany and C. Xie, "A complete review of incident detection algorithms and their deployment: What works and what doesn't," New England Transp. Consortium, Storrs, CT, Tech. Rep. 00-7, Feb. 2005.

[8] Z. Zheng, S. Ahn, D. Chen, and J. Laval, "Freeway traffic oscillations: Microscopic analysis of formations and propagations using wavelet transform," *Transp. Res. B: Methodol.*, vol. 45, no. 9, pp. 1378–1388, Nov. 2011.

[9] S. Thajchayapong and J. A. Barria, "Detection of lane-blocking on freeway segments using relative speed and lane changing trajectory," in *Proc. IEEE 9th Int. Conf. Intell. Transp. Syst. Telecommun.*, Oct. 2009, pp. 256–261.

[10] T. Nadeem, S. Dashtinezhad, C. Liao, and L. Iftode, "Traffic view: A scalable traffic monitoring system," in *Proc. IEEE Int. Conf. Mobile Data Manag.*, 2004, pp. 13–26.

[11] D. Helbing, *From Microscopic to Macroscopic Traffic Models*. Berlin-Heidelberg, Germany: Springer-Verlag, 1998.
[12] B. M. Williams and A. Guin, "Traffic management center use of incident detection algorithms: Findings of a nationwide survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 2, pp. 351–358, Jun. 2007.
[13] L. Iftode, S. Smaldone, M. Gerla, and J. Misener, "Active highways (position paper)," in *Proc. IEEE 19th Int. Symp. Personal, Indoor Mobile Radio Commun.*, Sep. 2008, pp. 1–5.
[14] P. Ferrari, "The reliability of the motorway transport system," *Transp. Res. B, Methodol.*, vol. 22, no. 4, pp. 291–310, Aug. 1988.
[15] J. Sheu, "A sequential detection approach to real-time freeway incident detection and characterization," *Eur. J. Oper. Res.*, vol. 157, no. 2, pp. 471–485, 2004.
[16] A. Chen, B. Khorashadi, C. Chuah, D. Ghosal, and M. Zhang, "Smoothing vehicular traffic flow using vehicular-based ad hoc networking and computing grid (VGRID)," *Proc. IEEE Intell. Transp. Syst. Conf.*, pp. 349–354, 2006.
[17] M. Abuelela, S. Olariu, and M. C. Weigle, "Notice: An architecture for the notification of traffic incidents," in *Proc. IEEE Veh. Technol. Conf.*, May 2008, pp. 3001–3005.
[18] G. K. Mitropoulos, I. S. Karanasiou, A. Hinsberger, F. Aguado-Agelet, H. Wieker, H. J. Hilt, S. Mammar, and G. Noecker, "Wireless local danger warning: Cooperative foresighted driving using intervehicle communication," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 539–553, Sep. 2010.
[19] M. Yongchang, M. Chowdhury, A. Sadek, and M. Jeihani, "Real-time highway traffic condition assessment framework using vehicle–infrastructure integration (VII) with artificial intelligence (AI)," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 615–627, Dec. 2009.
[20] L. Wischhof, A. Ebner, and H. Rohling, "Information dissemination in self-organizing intervehicle networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 90–101, Mar. 2005.
[21] M. S. Dougherty and M. R. Cobbett1, "Short-term interurban traffic forecasts using neural networks," *Int. J. Forecast.*, vol. 13, no. 1, pp. 21–31, Mar. 1997.
[22] R. Zito, G. Deste, and M. A. P. Taylor, "Global positioning systems in the time domain: How useful a tool for intelligent vehicle-highway systems?" *Transp. Res. C: Emerg. Technol.*, vol. 3, no. 4, pp. 193–209, Aug. 1995.
[23] K. Kiratiratanapruk and S. Siddhichai, "Vehicle detection and tracking for traffic monitoring system," in *Proc. IEEE TENCON*, Nov. 2006, pp. 1–4.
[24] R. L. Cheu and S. G. Ritchie, "Automated detection of lane-blocking freeway incidents using artificial neural networks," *Transp. Res. C: Emerg. Technol.*, vol. 3, no. 6, pp. 371–388, Dec. 1995.
[25] D. Srinivasan, X. Jin, and R. L. Cheu, "Evaluation of adaptive neural network models for freeway incident detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 1, pp. 1–11, Mar. 2004.
[26] P. C. Pendharkar, "A threshold varying bisection method for cost sensitive learning in neural networks," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 1456–1464, Feb. 2008.
[27] L. Leclercq, "Hybrid approaches to the solutions of the Lighthill–Whitham–Richards model," *Transp. Res. B: Methodol.*, vol. 41, no. 7, pp. 701–709, Aug. 2007.
[28] H. Adeli and A. Karim, *Wavelets in Intelligent Transportation Systems*. Hoboken, NJ: Wiley, 2005.
[29] K. Hornik, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, Jan. 1989.
[30] H. Dia and G. Rose, "Development and evaluation of neural network freeway incident detection models using field data," *Transp. Res. C: Emerg. Technol.*, vol. 5, no. 5, pp. 313–331, Oct. 1997.
[31] I. A. Basheer and M. Hajmeer, "Artificial neural networks: Fundamentals, computing, design, and application," *J. Microbiol. Methods*, vol. 43, no. 1, pp. 3–31, Dec. 2000.
[32] I. Maqsood, M. R. Khan, and A. Abraham, "An ensemble of neural networks for weather forecasting," *Neural Comput. Appl.*, vol. 13, no. 2, pp. 112–122, Jun. 2004.
[33] Vehicle information and communication system center, Accessed June 2009. [Online]. Available: http://www.vics.or.jp/english/vics/index.html
[34] O. Qing, R. Bertini, J. van Lint, and S. Hoogendoorn, "A theoretical framework for traffic speed estimation by fusing low-resolution probe vehicle data," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 747–756, Sep. 2011.
[35] N. An, J. Riihijarvi, and P. Mahonen, "Studying the delay performance of opportunistic communication in vanets with realistic mobility models," in *Proc. IEEE 69th Veh. Technol. Conf.*, 2009, pp. 1–5.
[36] B. S. Kerner, *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications and Theory*. New York: Springer-Verlag, 2004.
[37] Y. Guo, H. Zhang, X. Wang, and J. R. Cavallaro, "VLSI implementation of Mallat's fast discrete wavelet transform algorithm with reduced complexity," in *Proc. IEEE Global Telecommun. Conf.*, 2001, pp. 320–324.
[38] J. Ahmed, M. N. Jafri, J. Ahmad, and M. I. Khan, "Design and implementation of a neural network for real-time object tracking," in *Proc. World Acad. Sci., Eng. Technol.*, Jun. 2005, vol. 6, pp. 209–212.
[39] M. Pasquier and R. J. Oentaryo, "Learning to drive the human way: A step towards intelligent vehicles," *Int. J. Veh. Auton. Syst.*, vol. 6, no. 1/2, pp. 24–27, May 2008.

Suttipong Thajchayapong received the M.S. and B.S. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, and the Ph.D. degree in electrical and electronic engineering from the Imperial College London, London, U.K.

He is currently a Researcher with the National Electronic and Computer Technology Centre, National Science and Technology Development Agency, Pathumthani, Thailand. His research interests include intelligent transportation systems, with emphasis on vehicular traffic monitoring and inference, anomaly detection, and mobility and qualities of service in wireless networks.

Edgar S. Garcia-Treviño received the M.S. degree in electronic engineering from the Universidad de las Américas Puebla, Puebla, Mexico, in 2006. He is currently working toward the Ph.D. degree in electronic engineering with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K.

From 2001 to 2008, he was the Operations Manager with the Atmospheric Monitoring Network, Puebla State Environment and Natural Resources Agency. His research interests are focused on models for temporal data representation in the context of knowledge discovery and data mining.

Javier A. Barria (M'02) received the Ph.D. degree in electrical and electronic engineering and the M.B.A. degree from the Imperial College London, London, U.K., in 1994 and 1997, respectively.

He is currently a Reader with the Intelligent Systems and Networks Group, Department of Electrical and Electronic Engineering, Imperial College London. He has been the joint holder of several European Union Framework and U.K. Engineering and Physical Sciences Research Council project contracts, which are all concerned with aspects of communication systems design and management. His research interests include monitoring strategies for communication and transportation networks, network traffic modeling and forecasting using signal processing techniques, and distributed resource allocation in dynamic topology networks.

Dr. Barria is a Fellow of the Institution of Engineering and Technology and a Chartered Engineer in the U.K. He was a British Telecom Research Fellow from 2001 to 2002.