Ad Hoc Networks 36 (2016) 450–464

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

Clustering and sensing with decentralized detection in vehicular ad hoc networks☆

Andrea Gorrieri[a], Marco Martalò[a,b,*], Stefano Busanelli[c], Gianluigi Ferrari[a]

[a] Wireless Ad-hoc Sensor Network Laboratory, Department of Information Engineering, University of Parma, Parma, Italy
[b] E-Campus University, Novedrate (CO), Italy
[c] Guglielmo S.r.L., Barco di Bibbiano, Reggio Emilia, Italy

ARTICLE INFO

Article history:
Received 19 November 2014
Revised 21 May 2015
Accepted 27 May 2015
Available online 18 June 2015

Keywords:
Vehicular ad-hoc networks (VANETs)
Vehicular sensor networks (VSNs)
Clustering
Decentralized detection
Spatially constant phenomenon
Mobility

ABSTRACT

In the near future, vehicles will be more and more advanced sensing platforms: for instance, at least one smartphone (with several on-board sensors) is likely to be inside each vehicle. Smartphone-based inter-vehicle communications thus support the creation of vehicular sensor networks (VSNs). In this paper, we analyze the performance of clustered VSNs, where (hierarchical) decentralized detection schemes are used to estimate the status of an observed spatially constant phenomenon of interest. Clustering makes processing efficient and the architecture scalable. Our approach consists of the creation, during a *downlink* phase, of a clustered VSN topology through fast broadcast of control messages, started from a remote sink (e.g., in the cloud), through a novel clustering protocol, denoted as cluster-head election irresponsible forwarding (CEIF). This clustered VSN topology is then exploited, during an *uplink* phase, to collect sensed data from the vehicles and perform distributed detection. The performance of the proposed scheme is investigated considering mostly IEEE 802.11b (smartphone-based) as well as IEEE 802.11p (inter-vehicle) communications in both highway-like and urban-like scenarios. Our results highlight the existing trade-off between decision delay and energy efficiency. The proposed VSN-based distributed detection schemes have to cope with the "ephemeral" nature of clusters. Therefore, proper cluster maintenance strategies are needed to prolong the cluster lifetime and, as a consequence, the maximum amount of data which can be collected before clusters break. This leads to the concept of decentralized detection "on the move."

1. Introduction

In the last decade, commercial vehicles have witnessed an exponential growth of their sensing, computational, and communication capabilities. This huge improvement is enabling the implementation of a large number of innovative services and applications, including: safety, traffic management, smart navigation, environmental monitoring, etc. By exploiting their sensing and communication capabilities, the vehicles can cooperate to create so-called vehicular sensor networks (VSNs) [1]. VSNs have peculiar characteristics at various levels, from communication, networking, and data processing perspectives.

From a communication perspective, the vehicles continuously gather, process, and share location-relevant sensor data (e.g., road conditions, pollution, etc.). Information collection and dissemination can be performed using inter-vehicular

☆ This work has been partially presented at the 2011 International Conference on Intelligent Transportation Systems (ITST), St. Petersburg, Russia, August 2011.

* Corresponding author at: Wireless Ad-hoc Sensor Network Laboratory, Department of Information Engineering, University of Parma, Parma, Italy. Tel.: +39 521905759.

E-mail addresses: andrea.gorrieri@gmail.com (A. Gorrieri), marco.martalo@unipr.it (M. Martalò), stefano.busanelli@guglielmo.biz (S. Busanelli), gianluigi.ferrari@unipr.it (G. Ferrari).

http://dx.doi.org/10.1016/j.adhoc.2015.05.019
1570-8705/

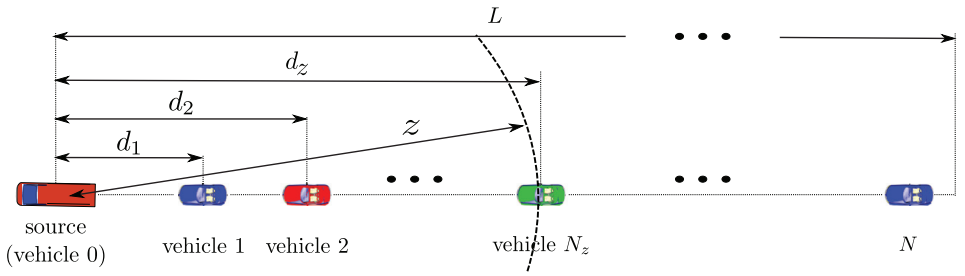

Fig. 1. Pictorial description of a linear VSN.

communications [2] and/or relying on the presence of roadside infrastructure [3]. Moreover, each vehicle is likely to contain at least a smartphone, which is itself a powerful sensing platform. In this context, cluster-based networking is an attractive solution to reduce network congestion and to simplify routing and data aggregation/dissemination [4].

An interesting approach to cluster-based vehicular communications can be found in [5], where communications are typically broadcast but, when possible, short-lived clusters are created in order to constitute a backbone, that can be used to support unicast communications. One of the strongest motivations for the design of cluster-based vehicular networks is provided by [6], where the authors show that, according to realistic mobility models, vehicular ad-hoc networks (VANETs) naturally evolve to clustered configurations. The advantages of clustering have also been exploited in the realm of decentralized detection [7], e.g., to determine optimum clustering and medium access control configurations. In particular, in [8] the authors provide a general framework for the computation of the probability of decision error when a spatially constant binary phenomenon is detected through a (possibly) multi-level sensor network.

The goal of this work is to present a decentralized detection scheme for data acquisition in clustered VSNs, which fits well with the requirements of on-demand detection applications. In particular, the proposed scheme might be used to determine, in a timely manner, if in a given city area (e.g., several blocks) there has been a critical situation (e.g., road congestion). A possible application of interest is the dissemination of this information to prevent other vehicles from running into this congested area (e.g., adaptive cruise control for congestion avoidance [9,10]). The proposed sensing and detection scheme foresees a two-phase communication mechanism. First of all, a *downlink phase* is triggered by a remote sink, with data collection duties, in order to form a clustered topology, constituted by ephemeral clusters (i.e., with limited lifetime) with associated cluster heads (CHs). The downlink phase is carried out through an innovative protocol, denoted as cluster-head election irresponsible forwarding (CEIF), which significantly improves the multihop probabilistic broadcast protocol, denoted as CIF, originally proposed in [11]. The so-formed clustered VSN is then used, during the (second) *uplink phase*, for data aggregation and/or local per-cluster fusion carried out at the CHs. The proposed scheme has been preliminarily presented in [12], where the basics of (i) the clustering protocol and (ii) the decentralized detection mechanism have been outlined together with preliminary results. While in [12] only static (steady-state) network conditions are considered, in this paper the performance of CEIF is analyzed by considering realistic dynamic ("on the move") conditions, in both highway-like and urban-like mobility scenarios. In particular, the performance of the proposed VSN clustered decentralized detection scheme is investigated considering mostly IEEE 802.11b communications between smartphones, as well as IEEE 802.11p between vehicles. Moreover, we also propose a novel reclustering procedure to be activated after ephemerals clusters break.

The rest of this paper is structured as follows. In Sections 2 and 3, preliminaries on the system and communication models, respectively, are provided. In Section 4, the decentralized detection mechanism is described. In Section 5, the performance of the proposed scheme is analyzed in a static scenario, i.e., under average steady-state conditions. In Section 6, we analyze the impact of mobility on the system performance in highway-like scenarios, from both clustering and sensing points of view. In Section 7, the system performance is investigated in a realistic urban-like scenario. Finally, concluding remarks are given in Section 8.

2. System model

Fig. 1 shows the linear network topology of reference for a VSN: N nodes are placed in a one-dimensional scenario. This is representative of a highway-like scenario—in Section 7, an urban-like scenario will be considered. Each node is uniquely identified by an index $i \in \{1, 2, \ldots, N\}$. The source node, denoted as node 0, is placed at the left end of the network. In order to derive the proposed clustering protocol (i.e., CEIF), we first consider steady-state conditions, i.e., a static network where nodes are positioned according to a one-dimensional Poisson point process with parameter ρ_s, where ρ_s is the linear vehicle spatial density (dimension: [veh/m])—the validity of this assumption is confirmed by empirical traffic data [13]. In Section 6, we will relax this assumption by analyzing more realistic VSNs with mobile nodes.

Each vehicle has a fixed transmission range, denoted as z (dimension: [m]), which depends on the transmit power and on the propagation model. In particular, the latter is assumed to be deterministic and the following models will be considered: Friis and Two Ray Ground [14]. Each vehicle is equipped with a global positioning system (GPS) receiver—namely, each on-board smartphone. As a consequence, each vehicle knows its own position at any given time—this is realistic in most vehicular conditions (but galleries). The maximum network length of the linear VSN is denoted as L (dimension: [m]), so that the number N of vehicles in $[0, L]$ can

be modeled as a Poisson random variable with parameter $\rho_s L$.

All vehicles observe a spatially constant phenomenon, i.e., a phenomenon whose status does not change from vehicle to vehicle along the road. For example, vehicles could monitor if the average vehicle spatial density on the road overcomes a critical threshold (i.e., there is traffic congestion): the VNS would declare that it does if it is declared by most of the vehicles on the road. This phenomenon is typical in applications such as adaptive cruise control, where the information on the local road conditions may be used to improve the behavior of vehicles approaching that road [9,10]. The observed binary phenomenon can be generally modeled as follows:

$$H = \begin{cases} H_0 & \text{with probability } p_0 \\ H_1 & \text{with probability } 1 - p_0 \end{cases}$$

where $p_0 \triangleq \mathbb{P}\{H = H_0\}$, being $\mathbb{P}\{\mathcal{A}\}$ the probability that the event $\mathcal{A}$ happens. The value H_0 can be interpreted as representative of a situation where the underlying physical phenomenon is, on average (along the road), below a given threshold. At the opposite, the value H_1 corresponds to the fact that the underlying physical phenomenon is, on average (along the road), above a given threshold. In the remainder of this paper, we will suppose, for notational simplicity, that $H_0 = 0$ and $H_1 = 1$.

3. Inter-vehicle communications and clustered VANET creation

In this section, we introduce the communication model behind the proposed VSN-based distributed detection scheme. First, a *downlink phase* is envisioned, where the sink broadcasts a query to all vehicles in the network, in order to obtain information about the phenomenon of interest. During this phase, the CEIF protocol, besides guaranteeing fast information dissemination, automatically creates a clustered VNS topology: each cluster has a single CH and all vehicles in the cluster communicate directly to it. After a clustered network topology has been generated, during the *uplink phase* in each cluster the data sensed by the vehicles (namely, by on-board smartphones) are sent to their corresponding CH where local (per cluster) fusion is performed. Then, the fused data are transferred from CHs to the sink, which takes the final decision on the status of the observed phenomenon. The uplink collection phase has to be completed before the clusters break down because of nodes' mobility [15].

3.1. Downlink phase

Before illustrating the basic operational principle of the CEIF protocol, it may be helpful to define the basic concepts of (i) multi-hop broadcasting and (ii) transmission domain (TD). In general, in multi-hop broadcasting scenarios a source node starts transmitting a packet which is directed to all other nodes in the network. This initial transmission is denoted as the 0th hop transmission, while the source itself identifies the so-called 0th TD. The packet transmitted by the source is then received by its neighbors, which rebroadcast the packet in order to propagate it to all nodes which are beyond the source node range—these neighbors constitute the 1st TD. In general, the group of nodes that rebroadcast in jth transmission hop is defined as TD_j. In the "basic" linear scenario described in Section 2, the number of TDs required to cover all the network is a random variable, denoted as N_{TD}, taking values in the set $\{\lceil L/z \rceil, \lceil L/z\ rceil + 1, \ldots, N\}$.

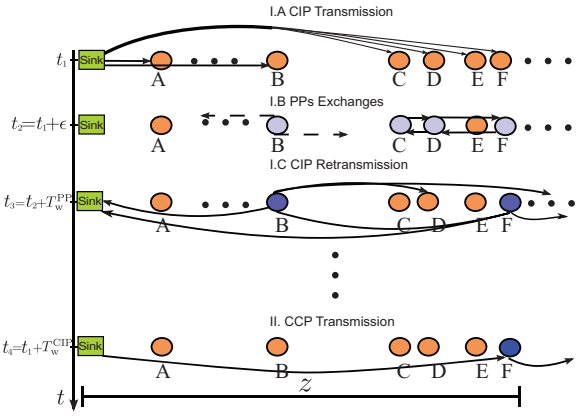

Fig. 2. Sequence of operation carried out by the CEIF protocol in TD_1.

The CEIF protocol has been designed in order to choose a single CH in every TD, thus creating a unique set of communicating CHs able to cover the entire area of interest. The nodes which are not designated as CHs become children of a CH, leading to the formation of clusters with similar dimensions. The CEIF protocol relies on 3 types of control packets: (i) Cluster Initialization Packet (CIP); (ii) Probe Packet (PP); (iii) Cluster Confirmation Packet (CCP). In Fig. 2, for illustrative purposes we show the message exchange in TD_1. More generally, CEIF involves two main operational subphases, which can be described as follows.

The first subphase requires the execution of three steps in every TD of the network (sequentially starting from TD_1) and terminates when the three basic steps have been completed in all TDs. We now describe the basic three steps with reference to the jth TD ($j = 1, \ldots, N_{\mathrm{TD}}$).

I.A At a generic instant t_1, a node of TD_{j-1} (the remote sink in the case of $j = 1$) sends a CIP with a transmit power $P_{\mathrm{t}}^{\mathrm{CIP}}$, containing its own position and a unique IDentification (ID). Upon the reception of a CIP from a node of TD_{j-1}, the receiver automatically becomes a member of TD_j.

I.B At $t_2 = t_1 + \epsilon$, where ϵ accounts for on-board processing time, the nodes of TD_j start competing to designate a CH. In particular, every node in TD_j elects itself as "candidate CH" according to the following heuristic probability assignment function [11]:

$$p = \exp\left\{-\frac{\rho_s(z-d)}{c}\right\} \tag{1}$$

where d is the distance between the sender and the receiver (dimension: [m]); c is a shaping coefficient (adimensional); $\rho_s z$ (dimension: [veh]) is the average number of vehicles within a transmission range. At this point, each candidate CH schedules the retransmission of a very short PP, bearing only the ID of the CIP and d, while the other nodes simply discard the CIP. The PPs are not forwarded and are transmitted with a high priority, in order

to reduce the overall latency, and with a low power which has been heuristically set to $P_t^{PP} = P_t^{CIP}/4$, in order to reduce network congestion and channel interference. Every candidate CH node sends a PP and then waits for a short interval, denoted as T_w^{PP} (dimension: [ms]): if, within this interval, it receives at least a PP containing a value of distance larger than its own, it stops forwarding and discards the packet; otherwise, it jumps to step I.C.

I.C At $t_3 = t_2 + T_w^{PP}$, the CIP is finally forwarded by the designated forwarding nodes of TD_j. The first subphase can be considered completed when the nodes of the last TD send their CIPs. Note that, since PPs are transmitted with a lower power, with respect to CIP, more than one candidate CH node may broadcast the CIP. For example, in Fig. 2 both nodes B and F rebroadcast the CIP.

The goal of the second subphase is to determine the actual CHs. It begins at epoch $t_4 = t_1 + T_w^{CIP}$, where $T_w^{CIP} \gg \overline{N}_{TD}(\epsilon + T_w^{PP})$ in order to guarantee that the second subphase starts after the average time required by the first subphase to complete. Initially, on the basis of its own information, every node shall elect its own CH. Note that a candidate CH can become aware of being the farthest candidate CH of its TD by simply listening to the CIP transmitted by the other candidate CHs of the same TD in the first subphase. If this is the case, the candidate CH elects itself as CH for its TD. At the same time (i.e., at t_4), the sink sends a CCP that shall be retransmitted only by the CHs, till the end of the network. By simply listening to the CCP, the remaining nodes can become aware of the identity of the true CHs.

According to the subphases of the downlink phase summarized above, the CEIF protocol can efficiently build, in a decentralized manner, a clustered topology, where each node elects its own CH without pursuing a common global consensus.

3.2. Uplink phase

The uplink phase exploits the clustered topology created during the downlink phase. More precisely, during the uplink phase, the data acquired by the N vehicles of the VSN are transmitted to the final remote sink in order to estimate the phenomenon status. Note that, unlike a static wireless sensor network, the created VSN can be used as long as its structure does not break down because of vehicle mobility. In other words, there is a maximum amount of data which can be collected. The impact of mobility will be investigated in detail in Section 6.

The observed signal at the ith vehicle can be expressed as

$$r_i = H \cdot s + w_i \qquad i = 1, \ldots, N \tag{2}$$

where $\{w_i\}$ are additive noise samples and s, which models the sensing quality, is considered as a deterministic quantity (the same for all vehicles). In particular, the parameter s is related to the sensor sensitivity. Assuming that the noise samples $\{w_i\}$ are independent random variables with the same Gaussian distribution $\mathcal{N}(0, \sigma^2)$, the common observation signal-to-noise ratio (SNR) at the vehicles is $SNR_{obs} \triangleq s^2/\sigma^2$ [8]. Each vehicle makes a decision comparing its observation r_i with a threshold value $\tau = s/2$ and computes a local decision $u_i = U(r_i - \tau)$, where $U(\cdot)$ is the unit step function. Note that a vehicle could transmit one single decision per packet or, by collecting consecutive phenomenon observations, it could transmit packets with a larger number of decisions. The selected strategy depends on the desired trade-off between data and overhead per transmitted packet. However, investigating this aspect goes beyond the scope of this paper.

Suppose that during the downlink phase the CEIF protocol has led to the creation of $N_c < N$ clusters. Each vehicle can communicate only with its local CH. Possible clustered topologies are shown in Fig. 3, according to the particular communication strategy toward the remote sink: (a) direct communications between the CHs and the sink (e.g., through an infrastructure-based network) and (b) multi-hop communications from the CHs to the sink.

We shortly comment on the remote sink. In Fig. 3, the sink is able to communicate with all the CHs, during the VSN lifetime, while they are moving. This may be a crucial issue in the presence of WiFi communications between the CHs and the sink. However, this limitation can be overcome assuming that the sink is a "logical" sink (e.g., a server in the cloud) and CHs can communicate with it through cellular communications (3G/4G). This, for instance, is the cross-network approach proposed in [16,17]. In the remainder of this paper, we do not further discuss this aspect, assuming that a CH, which has to communicate with the remote sink, can actually do it.

4. Fusion rule and probability of error

According to the theoretical framework presented in [8], in the presence of a spatially constant phenomenon a key performance indicator is the probability of decision error on the final phenomenon status estimate at the remote sink. This probability can be expressed as:

$$\overline{P}_e = \mathbb{P}\{\widehat{H} = H_1 | H_0\}\mathbb{P}\{H_0\} + \mathbb{P}\{\widehat{H} = H_0 | H_1\}\mathbb{P}\{H_1\}$$

where $\widehat{H}$ is the phenomenon estimate and the probabilities $\{\mathbb{P}\{\widehat{H} = H_\ell | H_m\}\}_{\ell,m=0}^{1}$ depend on the particular network topology (number of clusters and sensors per cluster) and the considered fusion rule. The fusion rule to be considered, either at a CH or at the remote sink, can be given by the following general expression:

$$\Phi(x_1, \ldots, x_M, k) \triangleq \begin{cases} 0 & \text{if } \sum_{m=1}^{M} x_m < k \\ 1 & \text{if } \sum_{m=1}^{M} x_m \geq k \end{cases} \tag{3}$$

where $x_1, \ldots, x_M$ are M binary data ($x_m \in \{0, 1\}$) to be fused together and k is the decision threshold. For even values of M, $k = M/2$; for odd values of M, $k = \lfloor M/2 \rfloor + 1$. Note that the binary data $\{x_m\}$ may be the data observed at the vehicles and to be fused together at the CH or the data, generated by the CHs, to be fused at the remote sink. The fusion rule in (3) is a majority fusion rule which guarantees a good performance in the presence of a phenomenon with equally likely statuses.

For a fixed value of the number of clusters, denoted as N_c, a clustered network topology can be described by the vector $\boldsymbol{\mathcal{D}} \triangleq (\mathcal{D}_c^{(1)}, \mathcal{D}_c^{(2)}, \ldots, \mathcal{D}_c^{(N_c)})$, where $\mathcal{D}_c^{(j)}$ is the number of nodes in the generic j-th cluster ($j = 1, \ldots, N_c$). As detailed in [8], if the network topology $\boldsymbol{\mathcal{D}}$ varies, the amount of information fused at each CH changes and, in turn, the probability

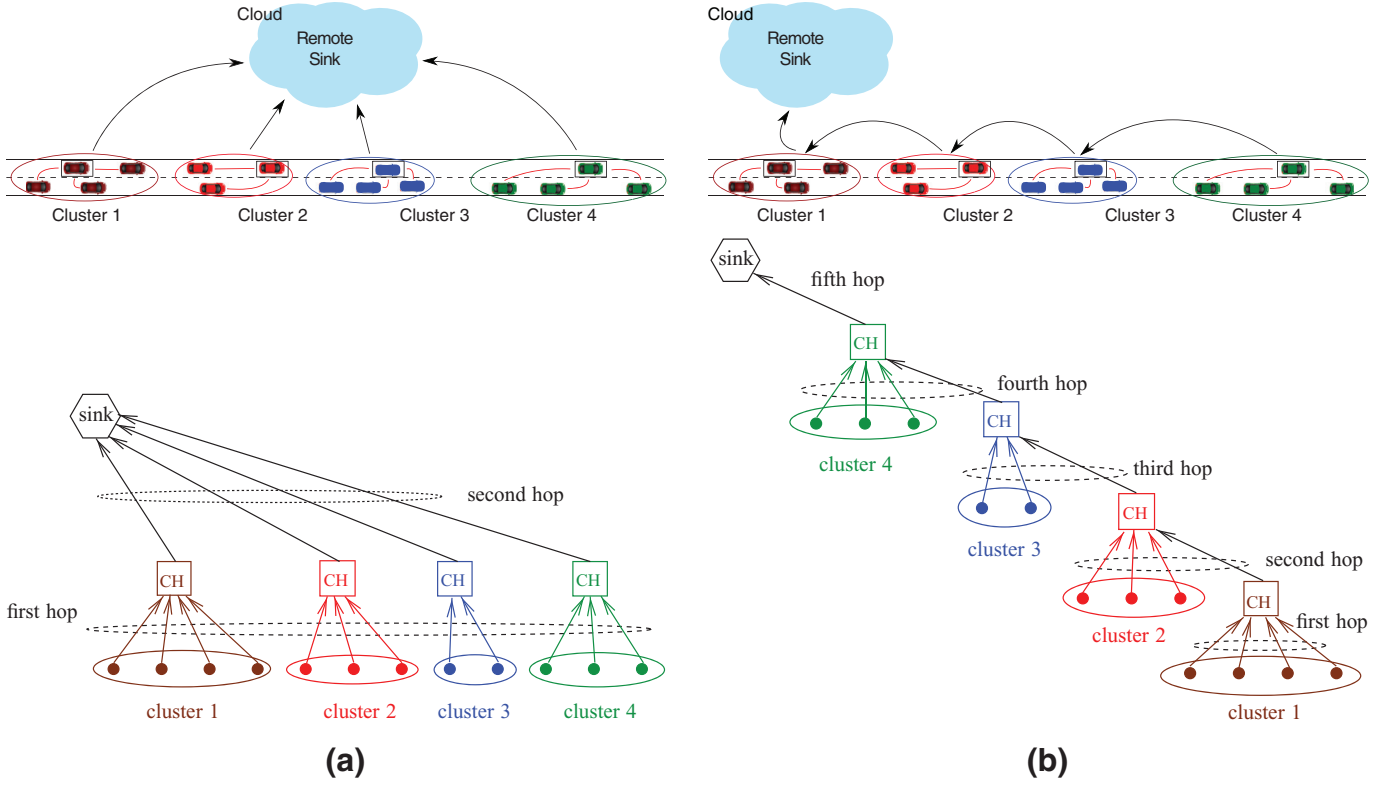

Fig. 3. Network topologies (upper part) and their logical representations (lower part): (a) direct communications between CHs and remote sink (b) multi-hop communications between CHs and remote sink.

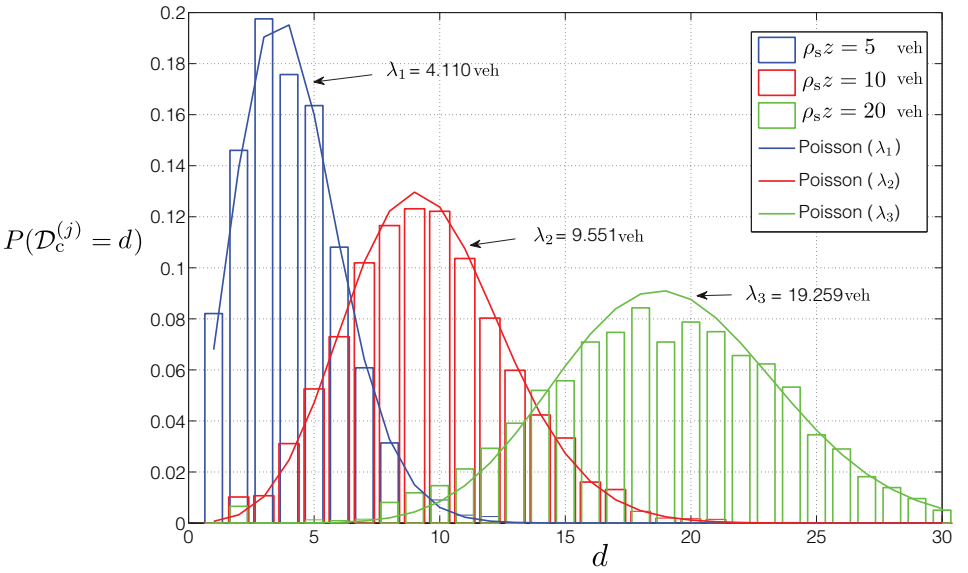

Fig. 4. PMF of the number of nodes per cluster. Various values of $\rho_s z$ are considered.

of error changes as well. Therefore, the probability of error is a function of $\boldsymbol{\mathcal{D}}$ and it is meaningful to compute the average, with respect to the Probability Mass Function (PMF) of $\boldsymbol{\mathcal{D}}$. In order to do this, the PMFs of $\mathcal{D}_c^{(j)}$, $j = 1, 2, \ldots, N_c$, and N_c are needed in (4). For the ease of simplicity, we set the value of the number N_c of clusters to its average $\overline{N}_{rmc}$. Under these assumptions, the average probability of decision error, with respect to the clustering configuration, can be computed as follows:

$$\overline{P}_e(\mathrm{SNR}_{obs}) = \mathbb{E}_{\boldsymbol{\mathcal{D}}|N_c=\overline{N}_c}[\overline{P}_e(\mathrm{SNR}_{obs}|\boldsymbol{\mathcal{D}}, N_c = \overline{N}_c)]. \quad (4)$$

5. Performance analysis in steady-state (static) scenarios

In this section, we first analyze the performance of the proposed CEIF protocol in a static scenario, i.e., without considering vehicles' mobility. This kind of scenario is representative of a steady-state mobile scenario in which the CH election procedure has been performed and clusters are static—this is a limiting case, in the presence of mobility, where all vehicles move at exactly the same speed. In such settings, the obtained performance is a benchmark, due to the fact that mobility breaks the clusters and, consequently, degrades the performance, as will be shown in Section 6.

5.1. Set-up

Simulations are carried out using the ns-3 simulator [18], with CEIF on top of an IEEE 802.11b wireless communication stack. The main parameters of the overall protocol (CEIF on top of IEEE 802.11b) are summarized in Table 1. We consider a linear network with length $L = 8z$ and $\rho_s = 0.02$ veh/m. The following values of $\rho_s z$ are considered: 5 veh ($z = 250$ m), 10 veh ($z = 500$ m), 15 veh ($z = 750$ m), 20 veh ($z = 1000$ m). Note that values of $\rho_s z$ smaller than 10 veh are representative of disconnected VSNs, whereas values larger than 10 veh are typical of (highly) connected networks.

Table 1
Main network simulation parameters for CEIF on top of IEEE 802.11b.

Packet size	100 bytes
CW_{min}	31
Carrier frequency	2.4 GHz
T_w^{PP}	10 ms
Data rate	1 Mbps
T_w^{CIP}	200 ms

We point out that all simulations have been carried out also considering IEEE 802.11p, which is specifically designed for VANETs, as MAC protocol [19]. However, the traffic load of the considered VSN is relatively small and, therefore, the MAC protocol has a negligible impact. This is in agreement with the results in [20], where a comparison between IEEE 802.11b-based and IEEE 802.11p-based VANETS, in terms of throughput, has been performed for small levels of traffic load (i.e., under the same conditions of our work). Therefore, we will focus on IEEE 802.11b as MAC protocol, which is more representative of VSNs based on on-board smartphones.

5.2. Results

The first parameter of interest for our analysis is the PMF of the number $\mathcal{D}_c^{(j)}$ of nodes in the generic jth cluster ($j = 1, \ldots, N_c$), which is needed to obtain the average probability of decision error (4). The final PMF is obtained by averaging over 500 simulation runs: for each of them a different network topology, corresponding to a specific configuration $\boldsymbol{D}$, is generated. In Fig. 4, the PMF is shown, considering various values of $\rho_s z$. As one can observe, the shape of the PMFs is the same, regardless of the value of $\rho_s z$. Despite a rigorous theoretical proof of this result is an open research problem, it can be observed that the curves have a Poisson-like shape. Using classical fitting tools, one obtains that the average

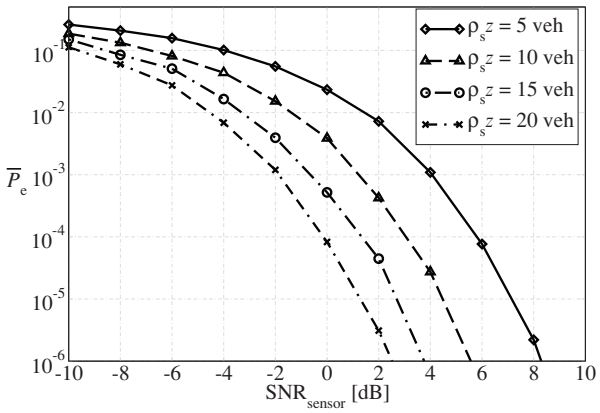

Fig. 5. Average probability of decision error at the remote sink, as a function of the vehicle observation SNR, in a scenario with CHs directly connected with the sink (scenario of Fig. 3(a)). Various values of $\rho_s z$ are considered.

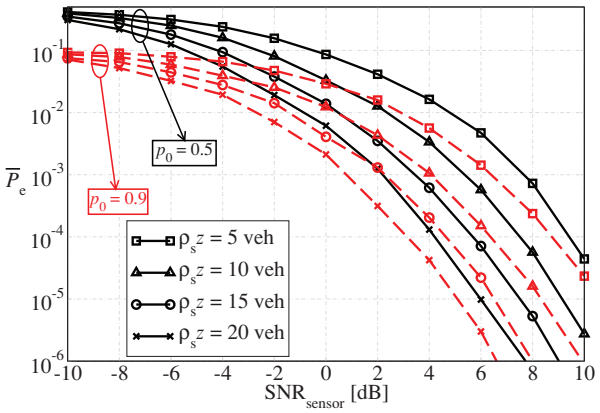

Fig. 6. Average probability of decision error at the remote sink, as a function of the vehicle observation SNR, in a scenario with CHs connected with the sink through multi-hop communications (scenario of Fig. 3(b)). Various values of $\rho_s z$ (5 veh, 10 veh, 15 veh, 20 veh) and of p_0 (0.5 and 0.9) are considered. The fusion rule (5) is used at the CHs.

value of the approximating Poisson distribution is 4.110 veh, 9.551 veh, and 19.259 veh for $\rho_s z$ equal to 5, 10, and 20 vehicles, respectively. As expected, for increasing values of $\rho_s z$, it is more likely to obtain larger cluster sizes, since there is a larger number of nodes in the transmission range of each other. In order to evaluate (4), the distribution of the number of clusters N_c is also needed. For the ease of simplicity, as mentioned earlier, we set this value to the average value $\overline{N}_c$ determined through simulations. More precisely, it follows that $\overline{N}_c = 7$ for $\rho_s z = 5$veh and $\overline{N}_c = 8$ for $\rho_s z$ equal to 10 veh, 15 veh, and 20 veh.

In Fig. 5, the probability of decision error is shown, as a function of the vehicle observation SNR, in a scenario with CHs directly connected with the sink (scenario in Fig. 3(a)). Different values of $\rho_s z$ are considered. As one can see, the larger the value of $\rho_s z$ is, the better is the performance. This has to be expected, since our simulations confirm the intuitive fact that a larger value of $\rho_s z$ corresponds to a larger average number of vehicles per cluster. In fact, it is well known that the performance of decentralized detection schemes, considering majority-like fusion rules, improves by using a larger amount of sensors [8].

Considering the inter-cluster multi-hop scenario in Fig. 3(b), majority-like fusion (according to (3)) can be carried out at each intermediate CH. In this case, however, the delay is much higher than in the scenario of Fig. 3(a), since intermediate communication and processing is needed before obtaining the final phenomenon status estimate at the remote sink. In general, since the signal processing time is negligible, with respect to the communication time, the delay can be approximated as $D \simeq n_h T_h$ (dimension: [s]), where T_h (dimension: [s]) is the time necessary to transfer a packet from one CH to the next one or to the sink (at each hierarchical level the packet size does not increase) and n_h is the maximum number of hops between the vehicles and the remote sink. For instance, in the scenario in Fig. 3(a) $n_h = 2$, whereas in the scenario in Fig. 3(b) $n_h = 4$ (more generally, $n_h = N_c + 1$).

From our results it turns out that the performance of the scheme in Fig. 3(b), with majority fusion at intermediate CHs, is poor, since only two binary decisions are fused at each CH. With $M = 2$, the decision threshold is $k = M/2 = 1$ and, according to the fusion rule in (3), a decision in favor of H_1 is taken if at least one of the two CHs is in favor of H_1. However, this decision rule is thus biased toward H_1 and this is detrimental in scenarios where the presence of the phenomenon (e.g., traffic congestion) is rare.

In order to evaluate the system performance when the phenomenon statuses are not equally likely, we consider the case of a rare event (i.e., $p_0 > p_1$). For this reason, we consider a majority-like fusion rule biased in favor of H_0, i.e., a decision in favor of H_1, at each intermediate CH in Fig. 3(b), is taken only if both CHs' decisions are in favor of H_1. This corresponds to setting $k = M = 2$ in (3), thus leading to the following fusion rule:

$$\Phi(x_1, x_2) \triangleq \begin{cases} 1 & \text{if } x_1 = x_2 = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Therefore, each CH locally decides applying the majority rule in (3) to the binary data coming from the vehicles in its own cluster. Then, the decision taken by the CH is fused with that coming from the CH in the previous hop using the modified rule in (5).

In Fig. 6, the probability of decision error is shown, as a function of the vehicle observation SNR, in the scenario of Fig. 3(b) considering the decision rule (5) at the CHs. Different values of $\rho_s z$ are considered. In a scenario with equal a priori probabilities of the phenomenon (with $p_0 = 0.5$), the performance drastically worsens with respect to the case with direct communications between the CHs and the sink (Fig. 5). This is due to the fact that, when the number of hops increases, the number of information fusions also increases and, therefore, the amount of information transferred across the network reduces, as already observed in [8]. Since, however, the fusion rule is biased in favor of H_0, the performance improves when the observed phenomenon is rare, e.g., in the case with $p_0 = 0.9$ and $p_1 = 1 - p_0 = 0.1$.

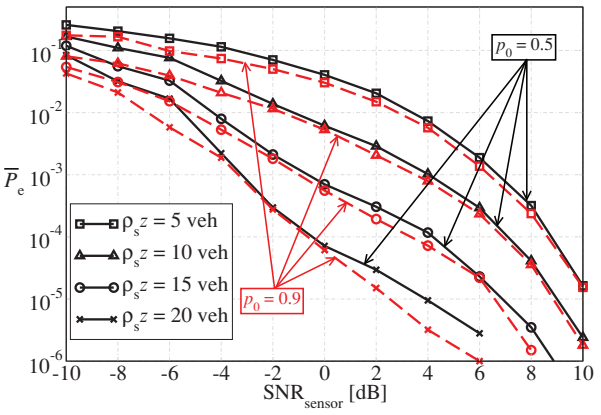

Fig. 7. Average probability of decision error at the remote sink, as a function of the vehicle observation SNR, for the scenario in Fig. 3(b). Various values of $\rho_s z$ and p_0 are considered. In all cases, the LLR-based fusion rule in (6) is used at each intermediate CH.

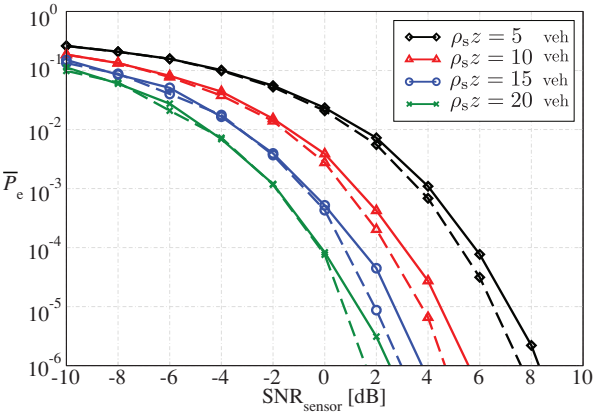

Fig. 8. Average error probability, as a function of the SNR, for various values of $\rho_s z$. The average error probability evaluated according to (4) (solid lines) is compared with the average probability of error assuming that all clusters have the same size $\overline{\mathcal{D}}_c$ (dashed lines).

5.3. Soft fusion

In order to improve the performance in VSNs with multi-hop topologies (Fig. 3(b)), we also consider a soft output-based fusion rule. Denote as $n_{d,1}^{(j)}$ the exact number of decisions in favor of H_1 at the jth cluster ($j = 1, \ldots, N_c$). Similarly, $n_{d,0}^{(j)} = d_c^{(j)} - n_{d,1}^{(j)}$ is the number of decisions in favor of H_0 in the jth cluster. At this point, we introduce the following logarithmic likelihood ratio (LLR):

$$\mathcal{L}_j \triangleq \ln \frac{n_{d,1}^{(j)}/d_c^{(j)}}{n_{d,0}^{(j)}/d_c^{(j)}} = \ln \frac{n_{d,1}^{(j)}}{n_{d,0}^{(j)}}$$

which corresponds to the logarithm of the ratio between the probability that the decision of a CH is in favor of H_1 and the probability that the decision of the CH is in favor of H_0. Obviously, $\mathcal{L}_j > 0$ if H_1 is more likely and $\mathcal{L}_j < 0$ if H_0 is more likely. Since $\forall j > 1$ each CH receives the LLR from the $(j-1)$th cluster, denoted as $\mathcal{L}_{j-1}^{\mathrm{UP}}$, the LLR generated at the jth cluster, to be passed to the elect CHs, can be expressed as follows:

$$\mathcal{L}_j^{\mathrm{UP}} = \mathcal{L}_{j-1}^{\mathrm{UP}} + \mathcal{L}_j.$$

Finally, the sink decides with the following rule[1]:

$$\widehat{H} = \begin{cases} H_0 & \text{if } \mathcal{L}_{N_c}^{\mathrm{UP}} < 0 \\ H_1 & \text{if } \mathcal{L}_{N_c}^{\mathrm{UP}} \geq 0. \end{cases} \tag{6}$$

In Fig. 7, the probability of decision error is shown, as a function of the vehicle observation SNR, in the scenario in Fig. 3(b), using the LLR-based fusion rule in (6). Comparing the results in Fig. 7 with those in Fig. 6, it can be concluded that the LLR-based fusion rule outperforms the majority fusion rule for both considered values of p_0 (0.5 and 0.9). This should be expected, since more information is transferred across the network—in fact, each CH transmits a LLR instead of a single bit. It is worth noting that this improved performance comes at the price of a higher energy consumption, due to the fact that the transmission of an LLR (even if quantized) requires a larger number of bits and, therefore, the energy consumption with soft fusion is higher than with majority fusion. However, depending on the chosen communication protocol, this energy penalty may be negligible. This is the case, for instance, with IEEE 802.11b communications.

5.4. Approximate performance analysis

A natural question arising at this point is the following. All simulation results in the figures presented in the previous subsections are obtained by averaging the probability of decision error with respect to the PMF of the number of nodes in each cluster. However, from Fig. 4 one can compute the average number of nodes per cluster, denoted as $\overline{\mathcal{D}}_c$, and assume that all clusters are composed by exactly $\overline{\mathcal{D}}_c$ vehicles. Therefore, it is interesting to understand the relation between $\overline{P}_e$ (as computed before) and the probability of error assuming that all clusters have the same number of nodes $\overline{\mathcal{D}}_c$. The latter can be computed according to the analytical framework proposed in [8] for the scenario of Fig. 3(a). In Fig. 8, the average error probability is shown, as a function of the observation SNR, considering the fusion rule in (3), for various values of $\rho_s z$. The performance of $\overline{P}_e$ (solid lines) is compared with the probability of error with clusters of the same size $\overline{\mathcal{D}}_c$ (dashed lines). One can observe that, for each value of $\rho_s z$, the gap between the two curves is limited. Moreover, the performances are trend-wise very close and, therefore, the "exact" average error probability can be accurately approximated, assuming that all clusters have the same size $\overline{\mathcal{D}}_c$. This result will be used in the next sections, where the performance in the presence of mobility will be analyzed. In particular, we will assume that, on average, all clusters will be composed by the same number of vehicles—this is meaningful in the presence of relatively uniform traffic conditions.

[1] Note that in the scheme in Fig. 3(b) the decision taken by the sink coincides with that taken by the last (hierarchically higher) CH.

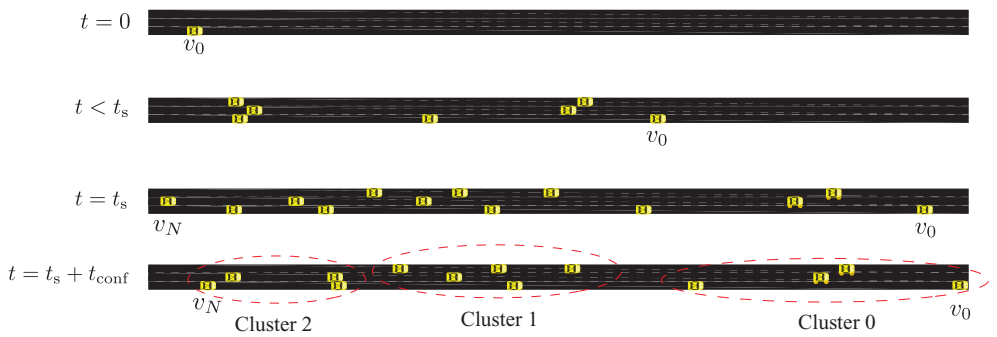

Fig. 9. Representative example of network evolution and clustering configuration in a mobile scenario.

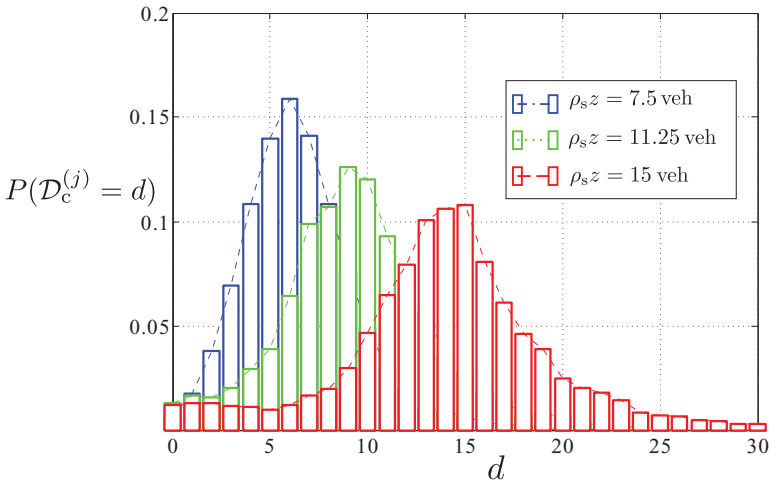

Fig. 10. PMF of $\mathcal{D}_c^{(j)}$ in the mobile scenario. Various values of $\rho_s z$ are considered.

6. Performance analysis: mobile scenario

In this section, we investigate the impact of vehicle mobility on the performance of the proposed vehicular decentralized detection scheme. In particular, all results refer to the 2-hop scenario in Fig. 3(a) with fusion rule given by (3). As anticipated in Section 3, this scenario is realistic since that next generation intelligent transportation systems are expected to be equipped with a heterogeneous mix of hardware and communication technologies [16,17].

6.1. Set-up

In order to analyze the impact of mobility, we have integrated the Simulation of Urban MObility (SUMO) traffic mobility suite [21] into the ns-3 simulator. SUMO implements a sophisticated car-following model which is an extension of those developed by Krauß [22,23].

Using SUMO, we have analyzed a highway-style scenario where the speed of each vehicle is a normally distributed random variable with mean μ_s (dimension: [m/s]) and deviation σ_s (dimension: [m/s]). According to the legal limits in Italian highways (namely, 35 m/s), we set $\mu_s = 35$ m/s and $\sigma_s = 4$ m/s. This means that half of the vehicles drive below the legal speed limit, but the other half of them exceed it—a realistic assumption in Italy. We consider a road with length $L = 40$ km with three lanes and an approximately constant vehicle spatial density ρ_s equal to 0.015 veh/m. The node range z is set so that the following values of $\rho_s z$ are obtained: 7.5 veh, 11.25 veh, and 15 veh.

6.2. Cluster formation

The network evolution, with an illustrative representation shown in Fig. 9, can be described as follows. At time instant $t = 0$, the first vehicle, denoted as v_0, enters the highway. Then, other vehicles (namely $v_i, i = 1, \ldots, N$) follow, one at a time, v_0 with their own speeds and abiding by the car-following model used by SUMO. At time instant t_s, the remote sink initializes the cluster formation and the last vehicle v_N starts the configuration process. From a simulation point of view, the starting time t_s is selected in order to make sure that all vehicles have entered the road. In particular, we consider $N = 200$ vehicles and the starting time is heuristically set to $t_s = 340$ s. The clusters are created at time $t_s + t_{conf}$, where t_{conf} is the time needed to create the clusters. In this scenario, nodes move overtaking each other and the clustering configuration changes as time goes by.

As a consistency check, in Fig. 10 we show the PMF of $\mathcal{D}_c^{(i)}$, $i = 1, 2, \ldots, \overline{N}_c$, in the mobile scenario introduced in the previous paragraph, at the end of the clustering procedure. The PMF is obtained by averaging over 500 trials where, in each

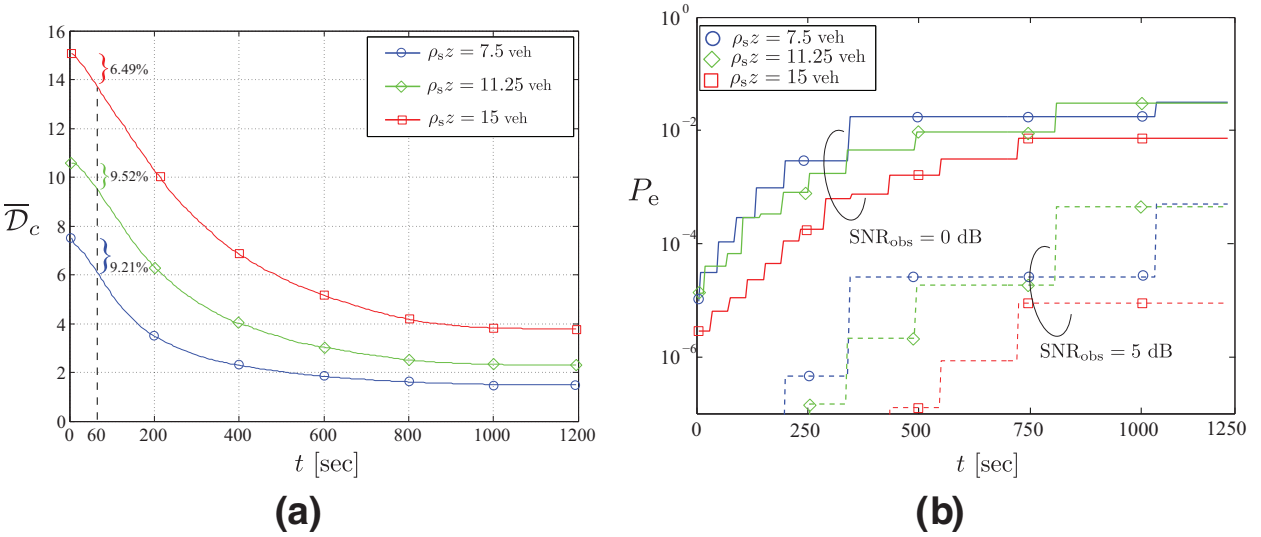

Fig. 11. Performance evolution, as a function of time, in terms of: (a) average number of nodes in a cluster $\overline{\mathcal{D}}_c$ and (b) average probability of error for two values of SNR$_{obs}$: 0 dB (solid lines) and 5 dB (dashed lines).

simulation run, a different mobility configuration is generated by SUMO. Note that the shape of the PMF is approximately the same of that in Fig. 4 for the static scenario. In particular, the cluster size is an increasing function of $\rho_s z$, since the number of nodes within the transmission range of each other increases. This result justifies the steady-state Poisson node distribution in the static scenario of Section 5.

6.3. Cluster evolution and network lifetime

As mentioned earlier in Section 5, the number of clusters N_c and the number of nodes in each cluster (i.e., the overall network topological structure $\boldsymbol{D} = \{\mathcal{D}_c^{(1)}, \ldots, \mathcal{D}_c^{(N_c)}\}$ have a direct impact on the average probability of decision error. In particular, mobility has a strong influence on the network topology and, as time goes by, the performance can heavily degrade, as clusters tend to "stretch," i.e., vehicles disconnect from their CHs.

In Fig. 11, the performance evolution is shown as a function of time in terms of: (a) average number of nodes in a cluster $\overline{\mathcal{D}}_c$ and (b) average probability of decision error. In case (b), two values of the observation SNR are considered: SNR$_{obs}$ = 0 dB (solid lines) or 5 dB (dashed lines). In case (a), one can observe that, as time evolves, the average number of nodes in a cluster decreases. This is due to the fact that vehicles have different speeds and, in a given cluster, a vehicle may overtake its CH, finally exiting from its transmission range. However, this process is relatively slow: in fact, after 1 min, the average reduction of the number of vehicles in a cluster is of 9.213%, 9.524%, and 6.487% when $\rho_s z$ is equal to 7.5 veh, 11.25 veh, and 15 veh, respectively. For $t \to +\infty$, $\overline{\mathcal{D}}_c$ is likely to tend to 1 (i.e., the CH remains alone), since all nodes may go outside the node range of the corresponding CH and it is very unlikely that two nodes have the same speed. This limiting behavior is not observed in Fig. 11, due to the fact that the road has a finite length. In order to determine the probability of error in case (b), the average number of nodes in a cluster is obtained by quantizing the curves of case (a). Moreover, our simulation results show that the average numbers of clusters $\overline{N}_c$ is 22, 15, and 13 for values of $\rho_s z$ equal to 7.5 veh, 11.25 veh, and 15 veh, respectively. It can be observed that the probability of error is a monotonically increasing function of the time and the case with $\rho_s z = 15$ veh guarantees the lowest value of $\overline{P}_e$ at any time. This is due to the fact that the larger the number of nodes per cluster, the lower the probability of decision error. Note that, for $t \simeq 750$ s, the case with $\rho_s z = 11.25$ veh corresponds to a higher value of $\overline{P}_e$ with respect to the case with $\rho_s z = 7.5$ veh. This is due to the fact that, even if the number of vehicles per cluster is quite similar in both cases, with $\rho_s z = 7.5$ veh the number of clusters is larger.

Since, as shown in Fig. 11(a), the average number of nodes per cluster decreases over time, it follows that clusters "break." Therefore, when $\overline{\mathcal{D}}_c$ reduces below a given threshold, the cluster creation procedure should be restarted to avoid that the probability of decision error becomes too high. Based on this motivation, the *network lifetime* is crucial and has a significant impact on the quantity of information that can be collected before the clustered network topology breaks down. In order to get more insights on the network lifetime, it is first necessary to clearly define this metric.

Various definitions of network lifetime have been proposed in the literature, depending on the application of interest. For instance, one can define the network lifetime as (i) the time interval (after clustered VANET creation) until the first cluster breaks or (ii) the time interval until a given Quality of Service (QoS), e.g., a given probability of error at the remote sink, is guaranteed. If the latter definition of network lifetime is considered, one can determine its value from the results in Fig. 11(b), by observing the time instant at which a target maximum probability of error is achieved. As an example, if the target probability of error (i.e., the chosen QoS indicator) is 10^{-4}, $\rho_s z = 15$ veh, and SNR$_{obs}$ = 0 dB, then the desired QoS is guaranteed until $t \simeq 250$ s. If, instead, the former definition of network lifetime is considered, then the *cluster lifetime* first needs to be defined. For example, one can

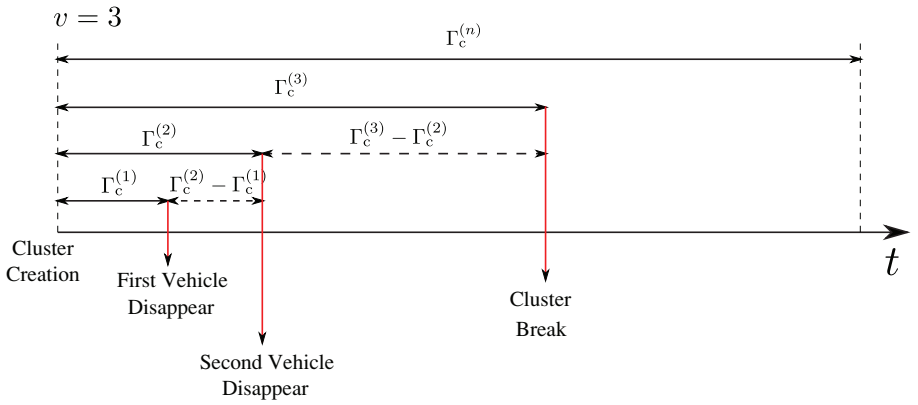

Fig. 12. Representative example of the cluster evolution process. The cluster lives until a number $v = 3$ of links breaks.

assume that a cluster breaks when a given percentage of its vehicles disappears, i.e., they are no longer connected to the CH (and, in turn, to the remote sink). We denote as $\Gamma_\ell^{(i)}$ the random variable representing the duration of the link between a CH and the ith of its children vehicles—for the jth cluster ($j = 1, \ldots, N_c$), $i \in \{1, \ldots, \mathcal{D}_c^{(j)}\}$. In [24], the authors prove that the PDF of $\Gamma_\ell^{(i)}$ has the following log-normal distribution:

$$f_{\Gamma_\ell^{(i)}}(\gamma) = \frac{1}{\gamma\sqrt{2\pi}\sigma_\Gamma} \exp\left[-\frac{(\ln\gamma - \mu_\Gamma)^2}{2\sigma_\Gamma^2}\right]$$

with μ_Γ (dimensions: [s]) and σ_Γ (dimensions: [s]) proper parameters which can be derived from experimental measurement campaigns.

Let us assume that a cluster breaks when the first of its vehicles disconnects. According to this assumption, the cluster duration can be given by the following expression:

$$\Gamma_c = \min\left\{\Gamma_\ell^{(1)}, \ldots, \Gamma_\ell^{(n)}\right\} \tag{7}$$

where n is the number of vehicles in a given cluster[2] (without considering the CH). Since each variable $\Gamma_\ell^{(i)}$ follows a log-normal distribution, it can be given the following expression:

$$\Gamma_\ell^{(i)} = e^{\mu_\Gamma + \sigma_\Gamma Z_i} \tag{8}$$

where $Z_i \sim \mathcal{N}(0, 1)$. By replacing (8) into (7), one obtains:

$$\Gamma_c = \min\left\{\Gamma_\ell^{(1)}, \ldots, \Gamma_\ell^{(n)}\right\} = e^{\mu_\Gamma + \sigma_\Gamma \min\{Z_1, \ldots, Z_n\}} = e^{\mu_\Gamma + \sigma_\Gamma Z} \tag{9}$$

where $Z \triangleq \min\{Z_1, \ldots, Z_n\} \sim \mathcal{N}(0, 1/n)$. This proves that, identifying the cluster death in correspondence to the instant of first vehicle disconnection from the CH, the PDF of the cluster lifetime is log-normal as well.

More generally, one can assume that a cluster lives until a given number v of links breaks (with $v > 1$). By denoting as $\Gamma_c^{(v)}$ ($v = 1, \ldots, n$) the random variable representing the time until v vehicles disconnect from the cluster, one can write:

$$\begin{aligned}
\Gamma_c^{(1)} &= \min\{\Gamma_\ell^{(1)}, \ldots, \Gamma_\ell^{(n)}\} = \Gamma_c \\
\Gamma_c^{(2)} &= \min\{\{\Gamma_\ell^{(1)}, \ldots, \Gamma_\ell^{(n)}\} \setminus \{\Gamma_c^{(1)}\}\} \\
&\vdots \\
\Gamma_c^{(v)} &= \min\{\{\Gamma_\ell^{(1)}, \ldots, \Gamma_\ell^{(n)}\} \setminus \{\Gamma_c^{(1)}, \ldots, \Gamma_c^{(v-1)}\}\} \\
&\vdots \\
\Gamma_c^{(n)} &= \min\{\{\Gamma_\ell^{(1)}, \ldots, \Gamma_\ell^{(n)}\} \setminus \{\Gamma_c^{(1)}, \ldots, \Gamma_c^{(n-1)}\}\}.
\end{aligned} \tag{10}$$

Referring to Fig. 12, one can express $\Gamma_c^{(v)}$ as follows:

$$\begin{aligned}
\Gamma_c^{(v)} &= \Gamma_c^{(1)} + (\Gamma_c^{(2)} - \Gamma_c^{(1)}) + (\Gamma_c^{(3)} - \Gamma_c^{(2)}) \\
&\quad + \cdots + (\Gamma_c^{(v)} - \Gamma_c^{(v-1)}).
\end{aligned} \tag{11}$$

Observing that $\Gamma_c^{(1)}$ is a log-normal random variable and $\{\Gamma_c^{(i)} - \Gamma_c^{(i-1)}\}_{i=2}^{v}$ are approximately log-normal [25], it follows that $\Gamma_c^{(v)}$ is approximately log-normal. The value of the parameters of the final log-normal distribution can be found in [26]. This means that even assuming that more than one vehicle has to disconnect to make the cluster die, the cluster duration still follows a log-normal distribution.

In Fig. 13, the empirical PDF of the cluster lifetime, obtained from 500 simulation runs, with $\rho_s z = 7.5$ veh, is compared with a log-normal distribution with optimized parameters. Two definitions of cluster lifetime are considered: $v = 1$ and v corresponding to 30% of the nodes in the cluster—note that similar results can be obtained for other values of $\rho_s z$. One can observe that the simulation results confirm the theoretical log-normal distribution. In particular, by minimizing the mean square error between the simulation-based results and the log-normal PDF one finds that the parameters of the log-normal distribution are: $\mu_\Gamma = 4.14$ s and $\sigma_\Gamma = 0.19$ s for $v = 1$; and $\mu_\Gamma = 4.86$ s and $\sigma_\Gamma = 0.15$ s for v corresponding to 30% of the number of nodes in the cluster.

Focusing on the results of Fig. 13, we remark that the cluster lifetime can be quite long. In fact, recalling that

$$\mathbb{E}\{\Gamma_c\} = \exp\left\{\mu_\Gamma + \frac{\sigma_\Gamma^2}{2}\right\}[\text{s}]$$

the cluster lifetime is approximately 64 s in the case of cluster death after the first link disconnection and 130 s in the case of cluster death after 30% of the links break.

[2] Note that, in the previous sections, we have denoted the number of vehicles in the j-th cluster as $\mathcal{D}_c^{(j)}$. However, for the ease of simplicity, in the following we denote the number of vehicles in the generic cluster as n.

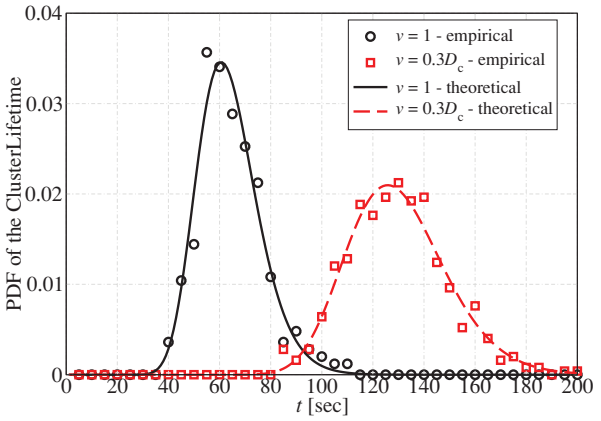

Fig. 13. Comparison of the empirical PDF of the cluster lifetime with the log-normal theoretical distribution of [24] for $\rho_s z = 7.5$ veh. Two definitions of cluster lifetime are considered: $v = 1$ and v corresponding to 30% of the nodes in the cluster.

Knowing the statistical distribution of the cluster lifetime, one can compute the total transferable amount of information before the cluster breaks, depending on the underlying communications/networking protocols. This goes behind the scope of this paper and represents an interesting research extension.

6.4. Cluster maintenance and reclustering

As outlined in Section 6.3, the cluster duration has a significant impact on the performance of the considered VSN. In order to prolong the overall network lifetime, different mechanisms can be envisioned to maintain the formed clusters and, eventually, to perform reclustering in the presence of vehicles leaving their own clusters. A possible solution is that all CHs periodically transmit, with period T_{ccp} (dimension: [s]), CCPs. A non-CH vehicle receiving a CCP checks if the CCP generator is at a distance shorter than $z/2$: if this is the case, the vehicle joins the new CH and, therefore, nodes exiting a cluster may have the possibility of joining a new one. The choice of the value of T_{ccp} is a tradeoff between a finer reclustering (and, therefore, a smaller information loss from vehicles) and a higher network overhead. One should note that, in the presence of mobility, some vehicles may leave the monitored area and portions of this area may even be uncovered. Since unconnected vehicles cannot participate in the data collection in the VSN, it is necessary to repeat, with period much longer than T_{ccp}, the entire CEIF clustering procedure described in Section 3.1. This aspect, however, has been neglected in our simulations, since the information loss caused by connectionless regions is negligible during our VSN operational lifetime.

In Fig. 14, $\overline{\mathcal{D}}_c$ is shown, as a function of time, in the presence of reclustering with $T_{ccp} = 100$ s. As already observed in Fig. 11(a), as time passes $\overline{\mathcal{D}}_c$ decreases due to mobility. However, when CCPs are generated, nodes outside their CHs' transmission ranges can now join a new cluster and the value $\overline{\mathcal{D}}_c$ thus suddenly increases. However, it is worth noting that, especially for the larger values of $\rho_s z$, the recovery procedure becomes less effective when time passes, since the value of $\overline{\mathcal{D}}_c$ immediately after each CCP generation is smaller and smaller. This is due to the fact that a larger value of $\rho_s z$ implies a larger cluster size and, therefore, the probability that a vehicle does not receive any CCP during reclustering increases.

7. Performance analysis in urban scenarios

In order to further investigate the impact of mobility on the performance of the proposed VSN, in this section we consider a urban-like vehicular scenario. One should note that the CEIF protocol and the corresponding clustering procedure have been designed for linear networks (e.g., highway-like) and are not optimized for bidimensional urban-like

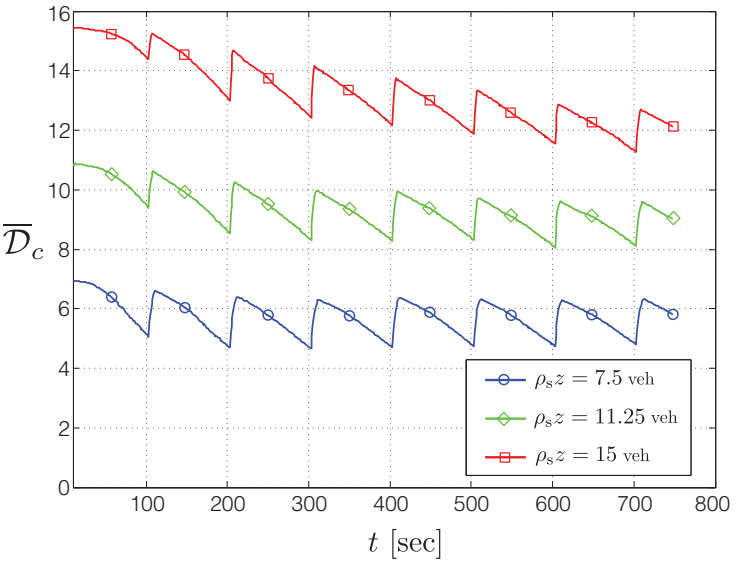

Fig. 14. $\overline{\mathcal{D}}_c$, as a function of time, in the presence of reclustering with $T_{ccp} = 100$ s.

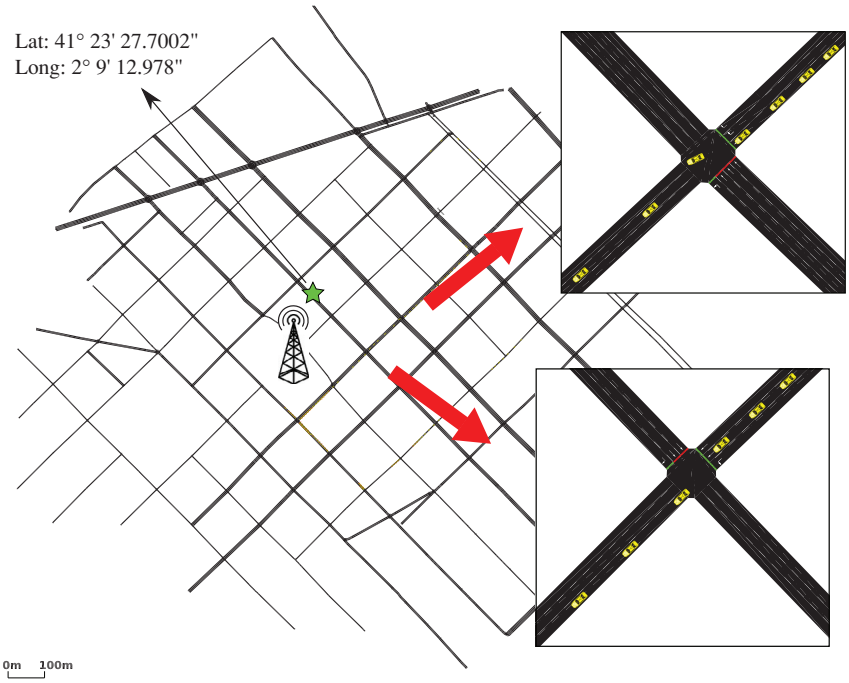

Fig. 15. Illustrative example of urban VNS scenario: portion of the city center of Barcelona imported into the SUMO mobility simulator.

scenarios. In this section, our focus is on a simulation-based investigation of the performance of CEIF in urban scenarios: the design of proper communication/networking protocols for such scenarios goes beyond the scope of this paper and may be the subject of future research.

The considered scenario is representative of the center of a large city with many road intersections, narrow roads, and single driving directions. Moreover, the nodes' speeds are highly heterogeneous: in fact, vehicles can move fast but they are constrained to abide by the traffic rules (priorities, traffic lights, etc.) forming queues and thus slowing down the overall vehicular traffic mobility. Since this is a highly dynamic scenario, more complex realistic mobility models are needed. To this end, we exploit the Open Street Maps (OSM) tool, which provides open and editable maps of real cities to be imported into SUMO [27]. As a representative vehicular scenario, we decided to simulate a portion of the city center of Barcelona, which is shown in Fig. 15. Vehicles move with a speed lower than or equal to the legal limits of the roads, stopping when needed, i.e., in correspondence to traffic lights and priorities. The node range z is set to 250 m and the number of considered vehicles is $N = 100$. At time instant $t_s = 550$ s the cluster formation procedure is triggered by the remote sink which, in this case, is represented by a base station placed in the center of the simulated area. As in the highway scenario considered in Section 6, the selected value of t_s guarantees that all vehicles enter the simulated area.

In Fig. 16, $\overline{\mathcal{D}}_c$ is shown, as a function of time, in the considered urban-like scenario depicted in Fig. 15. First, it can be noted that, because of mobility, $\overline{\mathcal{D}}_c$ is a decreasing function of time. However, the decaying is faster than in the highway-like scenario shown in Fig. 9: for instance, after 1 min the average reduction is of 65.962%. This is due to the fact that urban-like mobility is much more dynamic than highway-like mobility. As an example, traffic lights and priorities can be very detrimental for the cluster lifetime, as groups of vehicles may remain stopped in a queue thus leaving their own clusters. This also causes, unlike what happens in highway-like scenarios, the piecewise-constant trend of $\overline{\mathcal{D}}_c$. We can thus conclude that in the presence of urban-like mobility reclustering should be triggered more frequently.

8. Concluding remarks

In this paper, we have analyzed the performance of decentralized detection schemes for clustered VSNs. We have envisioned two phases: a downlink phase, during which a novel clustering broadcast protocol, denoted as CEIF, is employed; and an uplink phase, during which the vehicles perform, through the clustered topology, decentralized detection of a spatially constant phenomenon of interest. Different clustered topologies and fusion rules have been considered. The performance of the proposed VSN-based distributed detection schemes has been analyzed in terms of probability of error on the phenomenon estimate and network lifetime, considering different mobility models (namely, highway-like and urban-like). Unlike clustered sensor networks, where the clustering structure is a design aspect, the proposed vehicular distributed detection schemes exploit the natural formation of ephemeral vehicle clusters. Our results clearly show that the maximum amount of data collectible during the clustered VSN lifetime is more affected by the node mobility level than by the specific clustering structure. This suggest that "bursty" data collection strategies should be considered, together with proper (local) reclustering strategies. This is especially relevant in urban-like scenarios.

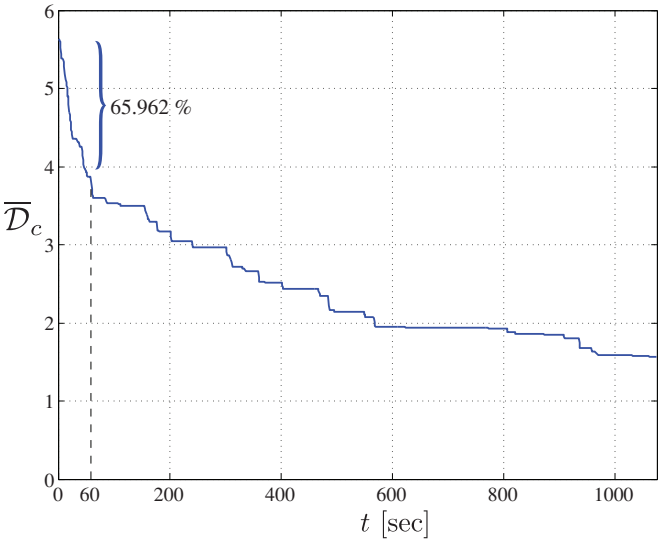

Fig. 16. $\overline{\mathcal{D}}_c$, as a function of time, in the considered urban-like scenario.

Acknowledgments

Part of the work of M.Martalò, S.Busanelli and G.Ferrari has been carried out under the one-year project "Cross-Network Effective Traffic Alerts Dissemination" (X-NETAD, Eureka Label E! 6252 [17]), sponsored by the Ministry of Foreign Affairs (Italy) and The Israeli Industry Center for R&D (Israel) under the "Israel-Italy Joint Innovation Program for Industrial, Scientific and Technological Cooperation in R&D," 2011.

The work of Andrea Gorrieri is partially supported by the SPINNER 2013.

References

[1] U. Lee, M. Gerla, A survey of urban vehicular sensing platforms, Elsevier Comput. Netw. 54 (4) (2010) 527–544, doi:10.1016/j.comnet.2009.07.011.
[2] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, A. Corradi, Dissemination and harvesting of urban data using vehicular sensing platforms, IEEE Trans. Veh. Technol. 58 (2) (2009) 882–901.
[3] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, S. Madden, CarTel: a distributed mobile sensor computing system, in: ACM International Conference on Embedded Networked Sensor Systems (SenSys), Boulder, CO, USA, 2006, pp. 125–138.
[4] C.R. Lin, M. Gerla, Adaptive clustering for mobile wireless networks, IEEE J. Select. Areas Commun. 15 (7) (1997) 1265–1275.
[5] L. Bononi, M.D. Felice, A cross layered MAC and clustering scheme for efficient broadcast in VANETs, in: IEEE Conference Mobile Ad-hoc and Sensor Systems (MASS), Montreal, Canada, 2007, pp. 1–8.
[6] M. Fiore, J. Härri, The networking shape of vehicular mobility, in: Proceedings of the ACM International Symposium on Mobile Ad-hoc Networking and Computing (MobiHoc), Hong Kong, China, 2008, pp. 261–272.
[7] M. Martalò, C. Buratti, G. Ferrari, R. Verdone, Decentralized detection in IEEE 802.15.4 wireless sensor networks, EURASIP J. Wireless Commun. Netw. 2010 (2010). Article ID 174063, 10 pages.
[8] G. Ferrari, M. Martalò, R. Pagliari, Decentralized detection in clustered sensor networks, IEEE Trans. Aerosp. Electron. Syst. 47 (2) (2011) 959–973.
[9] A. Kesting, M. Treiber, M. Schönhof, D. Helbing, Adaptive cruise control design for active congestion avoidance, Elsevier Transp. Res. Part C: Emerg. Technol. 16 (6) (2008) 668–683.
[10] A. Kesting, M. Schönhof, S. Lämmer, M. Treiber, D. Helbing, Decentralized approaches to adaptive traffic control, in: D. Helbing (Ed.), Managing Complexity: Insights, Concepts, Applications, Springer-Verlag, Heidelberg, Germany, 2008, pp. 189–199.
[11] S. Busanelli, G. Ferrari, S. Panichpapiboon, Cluster-based irresponsible forwarding, in: Tyrrhenian International Workshop on Digital Communications, Pula, Sardinia, Italy, 2009, pp. 59–68.
[12] S. Busanelli, M. Martalò, G. Ferrari, Clustered vehicular networks: Decentralized detection "on the move", in: Proceedings of International Conference on ITS Telecommunications (ITST), St. Petersburg, Russia, 2011, pp. 744–749.
[13] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, O.K. Tonguz, Routing in sparse vehicular ad hoc wireless networks, IEEE J. Select. Areas Commun. 25 (8) (2007) 1538–1556.
[14] K. Fall, K. Varadhan, The ns Manual, The VINT Project, 2010.
[15] G. Ferrari, O.K. Tonguz, Impact of mobility on the BER performance of ad hoc wireless networks, IEEE Trans. Vehicular Technol. 56 (2007) 271–286.
[16] S. Busanelli, F. Rebecchi, M. Picone, N. Iotti, G. Ferrari, Cross-network information dissemination in vehicular ad hoc networks (VANETs): experimental results from a smartphone-based testbed, MDPI Future Internet 5 (3) (2013) 398–428.
[17] Eureka project 6252 X-NETAD. http://www.eurekanetwork.org/project/-/id/6252.
[18] A. Wang, W. Jiang, Research of teaching on network course based on ns-3, in: International Workshop on Education Technology and Computer Science, Wuhan, China, vol. 2, 2009, pp. 629–632.
[19] F.A. Teixeira, V.F. Silva, J.L. Leoni, D.F. Macedo, J.M.S. Nogueira, Vehicular networks using the IEEE 802.11p standard: an experimental analysis, Elsevier Vehicular Commun. 1 (2) (2014) 91–96.
[20] B.E. Bilgin, V.C. Gungor, Performance comparison of IEEE 802.11p and IEEE 802.11b for vehicle-to-vehicle communications in highway, rural, and urban areas, Int. J. Vehicular Technol. 2013 (2013). Article ID 971684, 10 pages.
[21] F.K. Karnadi, Z.H. Mo, K. chan Lan, Rapid generation of realistic mobility models for VANET, in: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), Hong Kong, China, 2007, pp. 2506–2511.
[22] S. Krauß, Microscopic modeling of traffic flow: investigation of collision free vehicle dynamics, 1998 (Ph.D. thesis).
[23] D. Krajzewicz, Traffic simulation with SUMO—simulation of urban mobility, in: J. Barceló (Ed.), Fundamentals of Traffic Simulation, International Series in Operations Research & Management Science, vol. 145, Springer, New York, 2010, pp. 269–293.
[24] Y. Gongjun, S. Olariu, A probabilistic analysis of link duration in vehicular ad hoc networks, IEEE Trans. Intell. Transp. Syst. 12 (4) (2011) 1227–1236.
[25] S. Asmussen, L. Rojas-Nandayapa, Asymptotics of sums of lognormal random variables with Gaussian Copula, Stat. Probab. Lett. 78 (16) (2008) 2709–2714.

[26] N. Marlow, A normal limit theorem for power sums of independent normal random variables, Bell Syst. Tech. J. 46 (9) (1967) 2081–2089.
[27] M. Haklay, P. Weber, OpenStreetMap: user-generated street maps, IEEE Pervasive Comput. 7 (4) (2008) 12–18.

Andrea Gorrieri was born in Reggio Emilia, Italy, on October 27, 1987. He received the "Laurea" degree (3-year program, equivalent to a Bachelor) and the "Laurea" degree (3+2 program, equivalent to a Master) (summa cum laude) in Telecommunications Engineering in February 2010 and in July 2012, respectively, from the University of Parma, Italy. Since January 2013, he is a Ph.D. student in Information Technologies at the Department of Information Engineering of the University of Parma, Italy, supervised by Prof. Gianluigi Ferrari, and he is member of the Wireless Ad-hoc and Sensor Networks Laboratory (WASNLab).

Marco Martalò received his Ph.D. in Information Technologies from the University of Parma, Italy, in 2009. He was a Visiting Scholar at the EPFL, Switzerland (2007–2008). From May 2012, he is an Assistant Professor at the E-Campus University, Italy, and a Research Associate at the University of Parma, where he is member of the WASN Laboratory. His research interests are in the design of communication and signal processing algorithms for wireless systems and networks. Martalo is co-author of the book "Sensor Networks with IEEE 802.15.4 Systems: Distributed Processing, MAC, and Connectivity," and was a co-recipient of a "best student paper award" at IWWAN 2006, and also won, as part of the WASNLab team, the first prize award at the 2011 BSN Contest.

Stefano Busanelli received the "Laurea" degree (3 year program) in Telecommunications Engineering "summa cum laude" in December 2004. On December 2007, he received the "Laurea" degree (3+2 year program) in Telecommunications Engineering "summa cum laude." In March 2011, he got his Ph.D. in Information Technology from the University of Parma. From June 2008 to October 2008, he was an intern at Thales Communications (Colombes, France), collaborating with Dr. Christophe Le Martret and Dr. Isabelle Icart. From March 2011 to 2012, he was a Post-doc researcher at the Information Engineering Department (DII) of the University of Parma, under the supervision of Prof. Gianluigi Ferrari, working on wireless communications in vehicular ad hoc networks. From 2012, he is with Guglielmo S.r.l.

Gianluigi Ferrari (http://www.tlc.unipr.it/ferrari) is an Associate Professor of telecommunications at the University of Parma. Since 2006, he has been the Coordinator of the WASN Lab (http://wasnlab.tlc.unipr.it/) at the Department of Information Engineering. As of today, he has published more than 210 papers (journals and conferences) and book chapters, receiving a few awards. He co-authored 7 books. His research interests include wireless ad hoc and sensor networking, adaptive digital signal processing, and Internet of Things.