D4V: a peer-to-peer architecture for data dissemination in smartphone-based vehicular applications

Marco Picone, Michele Amoretti, Gianluigi Ferrari and Francesco Zanichelli

Department of Information Engineering, Università degli Studi di Parma, Italy

ABSTRACT

Vehicular data collection applications are emerging as an appealing technology to monitor urban areas, where a high concentration of connected vehicles with onboard sensors is a near future scenario. In this context, smartphones are, on one side, effective enablers of Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) applications and, on the other side, highly sophisticated sensing platforms. In this paper, we introduce an effective and efficient system, denoted as D4V, to disseminate vehicle-related information and sensed data using smartphones as V2I devices. D4V relies on a Peer-to-Peer (P2P) overlay scheme, denoted as Distributed Geographic Table (DGT), which unifies the concepts of physical and virtual neighborhoods in a scalable and robust infrastructure for application-level services. First, we investigate the discovery procedure of the DGT overlay network, through analytical and simulation results. Then, we present and discuss an extensive simulation-based performance evaluation (considering relevant performance indicators) of the D4V system, in a 4G wireless communication scenario. The simulation methodology combines DEUS (an application-level simulation tool for the study of large-scale systems) with ns-3 (a well-known network simulator, which takes into account lower layers), in order to provide a D4V proof-of-concept. The observed results show that D4V-based information sharing among vehicles allows to significantly reduce risks and nuisances (e.g., due to road defects and congestions).

INTRODUCTION

Driving safely, efficiently, and comfortably depends certainly on the vehicle status and on the driver behavior. However, a large number of external factors (e.g., traffic congestions, road defects, etc.) have a relevant impact and are difficult to predict without the support of ICT technologies. Among others, vehicular inter-networking has a prominent role (*Hartenstein & Laberteaux, 2009*), paving the way to several valuable applications, such as geocasting, mobile data sensing and storage, street-level traffic flow estimation, and others (*Lee & Gerla, 2010*). Vehicular inter-networking builds upon

Corresponding author
Michele Amoretti,
michele.amoretti@unipr.it

Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications, as well as on hybrid variants (*Hartenstein & Laberteaux, 2009*).

Recently, the Vehicular Sensor Network (VSN) research community has started investigating the possibility of using smartphones as V2V and V2I communication nodes, but also as portable sensing platforms (*Gerla & Kleinrock, 2011*). Smartphones are characterized by an ever improving technology—in terms of computational, networking, and storage capabilities—and good (practically ubiquitous) connectivity. Users often carry such powerful handheld devices in their cars, to take advantage of multimedia playback, navigation assistance, as well as Internet connectivity. Thus, in the near future, many vehicles may be exploited as mobile sensors to gather, process, and transmit data harvested along the roads in urban and extra-urban environments, potentially encompassing multiple types of information ranging from traffic/road conditions to pollution data and others.

As a matter of fact, until support for ad-hoc WiFi connectivity or for the new Wi-Fi-Direct standard (*Wi-Fi Alliance*) will be widespread, smartphone-based VSNs will require the presence of a communication infrastructure (e.g., 3G/4G cellular networks, WiMAX). Therefore, they will share the advantages of V2I schemes over V2V technologies, namely a better support in Commercial Off-The-Shelf (COTS) equipment, native long-range communication capabilities as well as support for broadcast or multicast communications (to the application level, at least). At the same time, cellular-network VSNs exhibit some disadvantages with respect to V2V schemes, such as: higher latency at short distances; local communication obtained only indirectly and by adding overhead; the need for service coverage; and the associated data traffic costs.

While hybrid communication schemes, i.e., combining V2V with V2I capabilities, would inherently provide the most effective and robust solution, we remark that purely infrastructure-based communication does not limit the application level of data dissemination and processing to a specific centralized architecture. In fact, a V2I infrastructure does not necessarily imply a centralized organization, which would inevitably lead to scalability issues—for example, to cope with the information requirements of thousands or millions of vehicles moving around in a large metropolitan area. While multiple distributed (e.g., hierarchical) subsystems can be deployed to achieve better scalability, a completely decentralized Peer-to-Peer (P2P) approach is more appealing. Initially exploited within V2V schemes (*Mahmoud & Olariu, 2007*), P2P approaches have been recently adopted also to implement decentralized Traffic Information Systems (TISs) (*Santa, Moragon & Gomez-Skarmeta, 2008*). In fact, P2P strategies allow responsibility decentralization, as well as computational and communication load balancing, which can be beneficial for smartphone-based VSNs (*Rybicki et al., 2007*).

In the context of P2P TISs, in this paper we introduce the D4V architecture, based on opportunistic mechanisms for the dissemination of data generated by vehicle sensors and drivers. D4V requires no dedicated hardware and leverages upon COTS and worldwide available devices (such as smartphones), rather than dedicated devices. Smartphones are usually affected by high energy consumption when using networks and GPS. However, this is not a real issue for vehicular applications, since smartphones will typically be connected

to an in-vehicle power source while using D4V. To the best of our knowledge, D4V is the only TIS providing, at the same time, massive scalability (because of its P2P nature), deployability (because of the light hardware requirements), and message configurability. With respect to the state of the art, D4V has a higher coverage percentage—the ratio between the number of peers that actually receive a specific message and the total number of those which should receive it—with less bandwidth requirements. D4V is based on a P2P overlay scheme denoted as Distributed Geographical Table (DGT) (*Picone, Amoretti & Zanichelli, 2011a*; *Picone, Amoretti & Zanichelli, 2011b*)—indeed, D4V stands for *DGT for VSNs*. The DGT overlay scheme represents a scalable and robust infrastructure for application-level services, and relies on the unification of the concepts of geographical and virtual neighborhoods (*Picone, Amoretti & Zanichelli, 2011a*). The DGT assumes that each peer knows its *Global Position* (GP)—which is reasonable, as nowadays every mobile device is equipped with a Global Positioning System (GPS).

The main contributions of this paper can be summarized as follows:

1. the introduction of the D4V system, with its opportunistic dissemination strategy for vehicular information and sensed data;
2. a sound analytical framework for performance evaluation of the DGT-based proactive neighbor localization protocol embedded in D4V; the DGT is introduced and partially analyzed in *Picone, Amoretti & Zanichelli (2011a)*, *Picone, Amoretti & Zanichelli (2011b)*;
3. the performance evaluation of the D4V, by means of multi-layer discrete event simulations, using realistic vehicle motion patterns and taking into account dangerous road stretches and traffic jams. In particular, we apply a recently proposed methodology (*Amoretti et al., 2013*) to integrate DEUS, an application-level simulation tool for the study of large-scale systems (*Amoretti, Agosti & Zanichelli, 2009*), with ns-3, a widely used simulation tool which takes into account lower layers (*ns-3 Development Team*).

The paper is organized as follows. Section 'Related Work' provides a summary of the state of the art about VSNs. Section 'Distributed Geographic Table' recalls the main DGT concepts. Section 'D4V System' illustrates the principles of the D4V system for traffic and sensed data dissemination. Section 'Mobility Model' illustrates the main aspects of the mobility model used to characterize realistic vehicular scenarios. In Sectiion 'Performance Evaluation', the results of the performance analysis of the proposed D4V system are presented: section 'DGT-based Proactive Localization' is dedicated to the analytical and simulation-based performance analysis of the DGT-based discovery protocol embedded in D4V; section 'D4V Performance Evaluation' analyzes the performance of D4V (both its DGT sublayer and its opportunistic dissemination strategy). Finally, section 'Conclusions' concludes the paper.

RELATED WORK

Several data dissemination strategies for VSNs have been proposed in the literature (*Lee & Gerla, 2010*). A fundamental issue is connectivity: different wireless access and communication methods have been evaluated, including Dedicated Short-Range Communication

(DSRC) (*Jiang et al., 2006*), WiMax/802.16e (*Han et al., 2008*), WLAN (*Hadaller et al., 2007*), as well as cellular systems (*Qureshi, Carlisle & Guttag, 2006*). The use of a cellular communication network reduces the problem of implementing a working TIS, but introduces, on the other hand, the issue of collecting and distributing data to interested users. In the following, we discuss some relevant research works related to VSNs, distinguishing between V2V and V2I approaches.

V2V approaches

Most V2V architectures rely on in-network aggregation mechanisms to improve communication efficiency by summarizing information that is exchanged between vehicles (*Dietzel et al., 2014*).

For example, in one of the earliest works in V2V communications, *Lee et al. (2006)* proposed MobEyes, a strategy for harvesting, aggregating, and distributing sensed data by means of proactive urban monitoring services provided by vehicles which continuously discover, maintain, and process information about events in an urban scenario. Messages and summaries are routed to vehicles in the proximity, to achieve common goals, such as providing police cars with the trajectories of specific target cars.

Later, *Meneguette et al. (2014)* proposed a network partition-aware geographical data dissemination protocol, which eliminates the broadcast storm and maximizes the data dissemination capability across network partitions with short delays and low overhead, at the expense of a high number of message transmitted. The same issue affects DOVE, proposed by *Yan, Zhang & Wang (2014)*.

V2V communication technology could mitigate traffic collisions and improve traffic congestion by exchanging basic safety information such as location, speed, and direction between vehicles within range of each other. Recently, *Xiang et al. (2015)* modeled vehicles' data preferences and explored the feasibility and benefits of incorporating these preferences into the design of safety data dissemination protocols. Furthermore, they designed PVCast, which assigns a higher transmission priority to packets that can satisfy a higher total data preferences of vehicles in the network by broadcasting. As a result, the differentiated transmission priorities of packets reduce contention and collision.

Another important aspect is the relationship between data dissemination performance and traffic congestion. *Du & Dao (2015)* recently developed analytical formulations to estimate information propagation time delay via a V2V communication network formed on a one-way or two-way road segment with multiple lanes. The proposed study carefully involves several critical communication and traffic flow features in reality, such as wireless communication interference, intermittent information transmission, and dynamic traffic flow. Moreover, this study elaborately analyzes the interactions between information and traffic flow under sparse and congested traffic flow conditions.

V2I approaches

Cellular networks are used in participatory VSN platforms, enabling applications such as ride quality monitoring, street-level traffic flow estimation, and proactive urban

surveillance (*Hull et al., 2006*; *Burke et al., 2006*; *Mohan, Padmanabhan & Ramjee, 2008*; *Mun et al., 2009*).

Initially, most participatory VSN platforms were based on the client/server paradigm, where all data generated by vehicles are stored in a central server (or server farm). *Hull et al. (2006)* pointed out that the major technical challenges of such a solution are mostly related with the huge amount of simultaneous updates and queries generated by user moves and requests (each car is a source of queries and regularly sends its own measurements).

For these reasons, researchers started investigating architectures based on the P2P paradigm, to build a distributed TIS where cars are not only consumers but also producers of information. *Rybicki et al. (2007)* with Peers on Wheels and, more recently, with PeerTIS (*Jedrzej et al., 2009*), proposed V2I architectures where participating cars are peers organized in a Distributed Hash Table (DHT). Roads are divided into road segments, each with a unique ID that is used as key in the DHT. The main idea is that each peer is responsible for a certain part of the ID space and, consequently, for a certain number of road segments. Up to now, one of the troubling issues is the fact that obtaining full information about planned and alternative routes is expensive in terms of bandwidth consumption. *Santa, Moragon & Gomez-Skarmeta (2008)* presented another P2P approach based on cellular networks and on the JXTA middleware (*Sun Microsystems, Inc, 2005*), to enable the transmission of information among vehicles and between vehicles and infrastructure, bounding the propagation of messages with respect to time and space (*Santa, Moragon & Gomez-Skarmeta, 2008*). A more recent location-aware P2P overlay scheme for smart traffic applications is Overdrive, developed by *Heep et al. (2013)*. Overdrive builds a geographical neighborhood for each peer, taking into account not only peers' positions, but also their speeds and directions. Traffic information is then disseminated by means of efficient flooding mechanisms.

DISTRIBUTED GEOGRAPHIC TABLE

A *structured* decentralized P2P overlay is characterized by a controlled overlay, shaped in a way that resources (or resource descriptors) are placed at appropriate locations (*Amoretti, 2009*). Moreover, a globally consistent protocol ensures that any peer can efficiently route a search to the peers that have the desired resource, even if the resource is extremely rare. Beyond basic routing correctness, two important topology constraints guarantee that (i) the maximum number of hops in any route (route length) is small, so that requests are fulfilled quickly, and (ii) the maximum number of neighbors of any peer (maximum node degree) is small, so that maintenance overhead is not excessive.

The DGT is a structured overlay scheme where each participant can efficiently retrieve information located near any chosen geographic position (*Picone, Amoretti & Zanichelli, 2011a*). In such a system, the responsibility for maintaining information about the position of active peers is distributed, for which a change in the set of participants causes minimal disruption. In the following, we recall the main DGT concepts using the P2P system notation introduced by *Aberer et al. (2005)*.

In a generic DGT overlay, the set of peers is called $\mathcal{P}$, each peer being characterized by a unique identifier $id \in \mathcal{I}$ (where $\mathcal{I}$ is the space of identifiers). The space of world's coordinates is denoted as $\mathcal{W}$ and $w \in \mathcal{W}$, $w = \langle latitude, longitude \rangle$ is a generic position. Thus, a peer $p \in \mathcal{P}$ may be identified by the pair $\langle id_p, w_p \rangle$, where $id_p \in \mathcal{I}$ and $w_p \in \mathcal{W}$. The distance d between two peers is defined as the actual geographic distance between their locations in the real world (also known as great-circle distance or orthodromic distance (*Longley et al., 2005*)). The *neighborhood* of a geographic location is the group of peers located inside a given region surrounding that location.

The main service provided by the DGT overlay is *request routing*, allowing to find available peers in a specific region, i.e., to determine the neighborhood of a generic global position $w \in \mathcal{W}$. Routing is a distributed process implemented as asynchronous message passing. By executing the *route* (p, w, a) operation, a peer forwards to another peer $p \in \mathcal{P}$ a request for the list of peers that peer p knows to be located in the region $a \in \mathcal{A}$, whose center is $w \in \mathcal{W}$. Thus, a routing strategy can be described as the following (possibly non-deterministic) function:

$$\mathcal{N} : \mathcal{P} \times \mathcal{W} \times \mathcal{A} \rightarrow 2^{\mathcal{P}} \tag{1}$$

which returns the neighborhood $\mathcal{N}(p, w, a)$, around the geographic position w and within region a, known by peer p.

The routing process is based on the evaluation of the region of interest centered in the target position. The idea is that each peer involved in the routing process selects, among its known neighbors, those that are likely to know a large number of peers located inside or close to the chosen region centered at the target point. If a contacted peer cannot find a match for the request, it does return the list of closest peers, taken from its routing table. This procedure can be used both to maintain the peer's local neighborhood $\mathcal{N}$ and to find available peers close to a generic target.

Regarding the local neighborhood, the general aim of the proposed approach is to provide accurate knowledge of peers that are close to the requesting one (starting from a set of neighbors provided by a bootstrap node, during the overlay network joining phase), together with a reduced set of remote peer references which will be used to forward long range geographic queries.[1] Whenever a single active peer in the system wants to contact other peers in its region (e.g., to provide or search for a service), it does not need to route additional and specific discovery messages to its neighbors (or to a supernode responsible for a specific zone) in order to find peers that are geographically close. Instead, it simply reads its neighbors' list, which is proactively filled with geographic neighbors.

Our peer neighborhood construction strategy has been inspired by Kademlia (*Maymounkov & Mazieres, 2002*), which is used, for example, in recent versions of the eMule client, as an alternative to the traditional eDonkey protocol (*eMule project*). Many of Kademlia's benefits result from its use of the XOR metrics to compute the distance between points in the resource identifier space. XOR is symmetric, allowing Kademlia participants to receive lookup queries from precisely the same distribution of peers contained in their routing tables, which are organized as sets of k-buckets. Every k-bucket is a list having

[1] Such an idea recalls Granovetter's theory of weak ties (*Granovetter, 1973*), stating that human society is formed by small complete graphs whose nodes are strongly connected (friends, colleagues, etc.). Such clusters are weakly connected between each other, e.g., a member of a group superficially knows a member of another group. The most important fact is that weak ties are those which make human society an *egalitarian small world network*, i.e., a giant cluster with small separation degree and no hubs (*Buchanan, 2003*).

up to k entries: in other words, each peer in the network has lists containing up to k peers, each list being associated with a given distance from the peer itself. In order to locate peers near a particular identifier, Kademlia uses a single routing algorithm from the beginning till the end. Peer neighborhood construction in DGT uses the geographic metric, instead of Kademlia's XOR metric. Each peer knows its GP, retrieved with a GPS or with other localization technologies (e.g., GSM cell-based localization), and knows a set of real neighbors organized in a specific structure, based on the distances of these neighbors from the peer's GP.

The *overlay network construction* is based on the process described in the following. Every peer maintains a set of *Geo-Buckets* (GBs), each one being a (regularly updated) list of known peers, sorted by their distances from the GP of the peer itself. GBs can be represented as K concentric circles, with increasing (application-specific) radii $\{R_i\}_{i=1}^K$ and thicknesses $\{r_i\}_{i=1}^K$, with $R_i = \sum_{j=1}^i r_j$. If there is a known peer whose distance is larger than the radius of the outmost circle R_K, it is inserted in another list which contains the peers located outside the region covered by the circle model.

Each peer in the GB set is characterized by (i) an *Identifier* (ID) which uniquely identifies the peer within the DGT; (ii) a GP, with latitude and longitude values retrieved by means of a GPS receiver or other positioning systems; (iii) a *Contact Address* allowing the identification of the peer in the Internet; (iv) and a *number of known peers* used to compare two peers which have the same distance. Moreover, each peer maintains a set of message types which it is interested on.

When a new peer wants to join the network, it sends a join request, together with its GP, to a *bootstrap node*, which returns a list of references to peers that are geographically close to the peer itself. It is important to emphasize that this information is not updated: referenced peers may have moved away from their initial locations. It is up to the joining peer to check for the actual availability of listed peers. Such an operation is performed when the peer first joins, but also when the peer finds itself to be almost or completely isolated. In these situations (that typically arise when peers enter low density regions), the peer may send a new join request to the bootstrap node, in order to obtain a list of recently connected peers which may become neighbors.

The main procedure used during *peer discovery* is *FIND_NODES(GP)*, which returns the β peers that are nearest to the specified GP. Peer p keeps its neighborhood awareness up-to-date by periodically applying *FIND_NODES* to its own global position GP_p. Such a procedure (with any target GP) may also be executed upon request from another peer.

Peer p searches in the GB related to the requested GP. The final objective of the peer lookup process is to find the $\alpha \leq K$ peers that are nearest to the selected GP, including newly connected peers, as well as mobile peers that have entered the visibility zone. The lookup initiator starts by picking α peers from its closest non-empty GB—or, if that bucket has less than α entries, it just takes the α closest peers, by extending the lookup to all its GBs. Such a peer set is denoted as $\mathcal{C}_i^{(0)} = \{p_i^{(0)}(1), \ldots, p_i^{(0)}(\alpha)\}_{i=1}^K$. The initiator sends parallel *FIND_NODES* requests, using its *GP* as target, to the α peers in $\mathcal{C}_i^{(0)}$. Each interrogated peer replies indicating β references. The initiator sorts the result list

according to the distance from the target position, then picks up α peers that it has not yet queried and re-sends the *FIND_NODES* request (with the same target) to them. If a round of *FIND_NODES* fails to return a peer closer than the closest one already known, the initiator re-sends the *FIND_NODES* to the K closest peers not already queried. The lookup terminates when the initiator has obtained responses from the K closest peers, or after f cycles, each jth cycle generating an updated set $\mathcal{C}_i^{(j)}$ of nearest neighbors. Thus, the number of sent *FIND_NODES(GP)* messages is $f \cdot \alpha + K$ and depends on the peer spatial density in the region of interest. A peer is allowed to run a new lookup procedure only if the previous one is completed, in order to reduce the number of exchanged messages and to avoid the overlapping of the same type of operations.

The general idea is that soon after the bootstrap or when neighbor peers are highly dynamic, the discovery process period may be very short and may increase when the knowledge among active peers becomes sufficiently stable.

Any active peer in the network can change its geographic position for many reasons (e.g., the user may be walking, driving, etc.). In order to preserve the consistency of the DGT, each peer has to periodically schedule a maintenance procedure which compensates the network topological changes. The practical usability of a DGT critically depends on the messaging and computational overhead introduced by such a maintenance procedure, whose features and frequency of execution are application-dependent.

When an active peer in the network changes its geographic position, it has to send updates of its GP to its neighbors, in order to make their knowledge more accurate. To avoid excessive bandwidth consumption, every peer communicates its position update to neighbors only if the displacement (with respect to the last communicated position) is higher than ϵ (dimension: (km)). If a peer receives a neighbor's update indicating that the new position of the latter is out of its region of interest, the neighbor's reference is removed from the appropriate GB and a *REMOVE* message is sent to the peer.

The DGT allows peers to have accurate knowledge of geographically close neighbors and a limited view of the outer world. However, whenever necessary and with limited incremental computational and transmission costs, a peer can find new peers which are entering the target region. The described P2P localization scheme represents the core layer of the proposed D4V system, which can discover and inform drivers who may be interested in specific traffic messages or data acquired by vehicle sensors.

D4V SYSTEM

Most vehicular network safety applications need information from a very limited geographic region around the vehicle's current position. This may not be the case for driving comfort applications, such as traffic intensity or traffic jam monitoring, as well as parking discovery (*Caliskan, Graupner & Mauve, 2006*) or guidance systems, which distribute information about the traffic status of the entire city or of those regions where the car is located or moving to.

The goal of the D4V system, based on the DGT overlay scheme recalled in Section 'Distributed Geographic Table', is to provide a reliable and scalable solution for
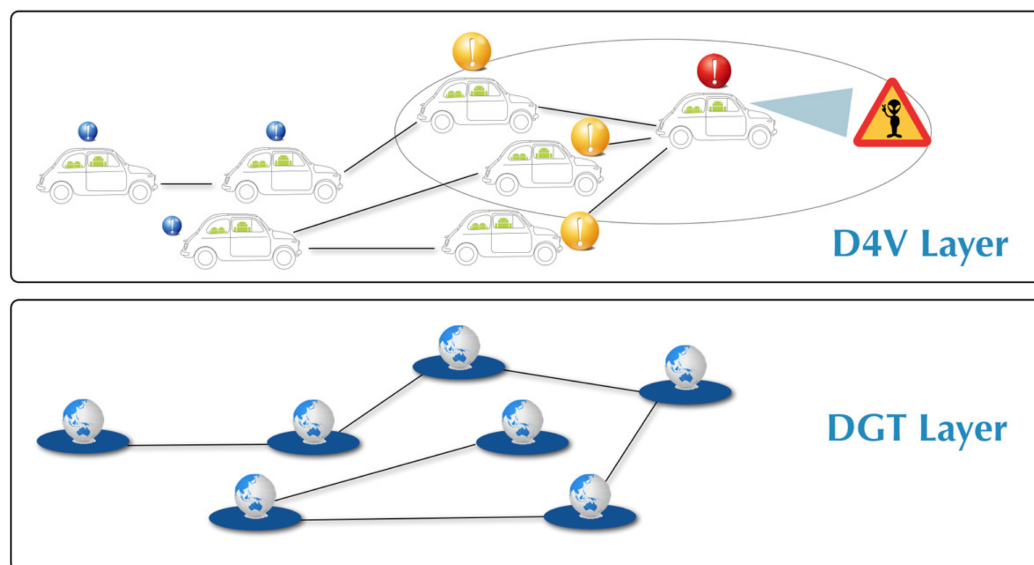
Figure 1 D4V and DGT layers.

disseminating—in an opportunistic way—information coming from driver's inputs or vehicle sensors, such as, for example, active shock absorber, cameras, engine, and temperature sensors. In Fig. 1, an illustrative representation of the D4V system, built on top of DGT, is shown.

Generally speaking, distributing information over long ranges in vehicular applications is a very challenging task in terms of how to gather, transport, and aggregate such information. The reference scenario of this paper is related to the case where the vehicular network and the user are interfaced uniquely through a mobile device. The information that enriches the knowledge base of the car is collected from internal and external data sources, namely vehicle or roadside infrastructure sensors. The on-board intelligence of the car extends, maintains, and disseminates such an information by creating a local view of the car surroundings.

In the literature, different techniques for content dissemination in VSNs are described, such as flooding and geocasting (*Tonguz et al., 2007*; *Bronsted & Kristensen, 2006*), request/reply (*Zhao & Cao, 2007*; *Wegener et al., 2007*), broadcasting, sharing (*Seskar et al., 1992*), and beaconing (*Xu & Barth, 2006*; *Fujiki et al., 2007*). In the D4V system, we combine the DGT scheme with the opportunistic and spatiotemporal dissemination approach proposed by *Leontiadis & Mascolo (2007)*, which is based on the publish/subscribe paradigm and allows message distribution to all interested receivers in a given region, by keeping messages alive in that region for a specific period of time. Owing to its properties, we believe that such an integrated solution fits well with a very dynamic scenario, where users can easily and frequently change their subscription interests according to their planned paths, the current season, their city neighborhood, and other parameters.

The basic D4V message is composed by: a *type*, for the notification category (for example, the class of traffic events or sensor data); a *location*, associated with the information; an *Event Range* (ER), which represents the region that the notification should reach; an *expiration time* of the event; and a message *payload* containing—whenever necessary—additional and detailed information about the event.

Different types of messages can thus be distributed by means of the same dissemination protocol. It is possible to create, for example, a message to warn approaching users about a traffic queue or a dangerous situation, to distribute data extracted from the different sensors of the vehicle or to notify other users about a free space in a parking area. Each user selects the list of message types he/she is interested on, and adds it to his/her peer descriptor, thus allowing other peers to send only appropriate messages, according to the receiver's preferences. When a new message is generated, the publisher picks up from its GBs the closest known peers of the DGT overlay, within the Event Range, that are interested in the particular information type (by reading the peer descriptor) and sends them the new message. Such an optimization can be obtained at the expense of a small overhead, due to the inclusion in the message of the list of previous recipients. When a notification is received, D4V checks if it matches the user interests or not (in the presence of dynamic subscription) or if it is already known. In the case of a new information, the peer adds it to its knowledge base, and distributes it again to known interested peers.

When a peer receives the references about a new peer in its region of interest, it checks if in its knowledge base there are notifications not yet expired that may be useful for that new peer. If the latter needs such an information (an hash-based comparison is performed), the peer provides it. During such a dissemination process, it is necessary to check if some messages have expired, and, consequently, to remove them and their references from the peer's knowledge base, thus avoiding the distribution of obsolete notifications.

MOBILITY MODEL

The mobility model is one of the fundamental elements for realistic performance evaluation of simulated V2V and V2I networking applications. In our work, we take into account some ideas proposed by *Harri, Filali & Bonnet (2005)* and *Fiore et al. (2009)*, about the key features that should be included in a vehicular mobility simulator, in order to obtain realistic motion patterns. Moreover, our model partially follows the approach of *Zhou, Xu & Gerla (2004)*, where the key idea is to use *Switch Stations* (SSs), connected via virtual tracks, to model the dynamics of vehicle and group mobility. For example, our simulative analysis considers a square area around the city of Parma (Italy), with 20 SSs inside and outside the city district: a Google Maps based representation of Parma is shown in Fig. 2. Stations are connected to each other through virtual paths, which have one lane for every direction, speed limitations associated with the street category, and a specific road density limit to model vehicles' speeds in traffic jam conditions. When a new car joins the network, it first associates with a random SS; then, it selects a new SS and starts moving along the connection path between the two SSs. Such a procedure is repeated every time the car reaches a new SS and has to decide its next destination. Each SS has

Figure 2 Example of simulated DGT-based VSN (in the city of Parma, Italy).

an attraction/repulsion value which influences the user's choice for the next destination station. This value may be the same for each path in order to allow a random trip selection.

A set of parameters is associated with each car, thus affecting macroscopic and microscopic aspects of traffic circulation, like street and highway limitations (i.e., some types of vehicles are forbidden on particular paths), as well as acceleration, deceleration, and speed constraints. We model different external events which may happen during the traffic simulation and alter drivers' behavior, such as: accidents, temporary road works, or bad road surface conditions due to ice, snow or potholes. We assume that these events can be detected by vehicle sensors. Drivers do not only interact with obstacles, but also adapt their behaviors according to their knowledge about surroundings. For example, they may try to change their paths if they are informed about a traffic jam or an accident slowing or blocking, and they reduce their speed in proximity of locations characterized by bad surface conditions.

We consider a microscopic flow model, where mobility parameters of a specific car are described with respect to other cars. Several approaches take into account, for example, the presence of nearby vehicles when modeling vehicle speed (e.g., Fluid Traffic Model (FTM) (*Seskar et al., 1992*; *Krauss, Wagner & Gawron, 1997*), and Intelligent Driver Model (IDM) *Trieber, Hennecke & Helbing, 2000*)). In particular, FTM is the most accurate for our scenario, with different speed limits for different virtual paths and low computational

requirements. In FTM, vehicle speed is a monotonically decreasing function of the vehicular spatial density, forcing lower values when the traffic congestion reaches a critical point. In our case, the desired speed of a vehicle moving along the points of the ith path is computed according to the following equation:

$$v^{\text{des}} = \max\left\{ v_{\min}, v^{(i)}_{\max}\left(1 - \frac{k}{k_{\text{jam}}}\right)\right\} \tag{2}$$

where: $v_{\min}$ is the minimum vehicle speed and depends on the vehicle's characteristics (dimension: (km/h)); $v^{(i)}_{\max}$ is the speed limit of the ith path (dimension: (km/h)); k is the current vehicular spatial density of the road (dimension: [vehicles/km]), given by n/l (n represents the number of vehicles on the road and l is its length in km), and k_{jam} is the vehicular spatial density (dimension: (vehicles/km)) in correspondence to which a traffic jam is detected.

As mentioned before, we also want to model the behavior of a driver in proximity of a road point with bad surface conditions. The idea is that a conscientious driver, knowing that along his/her road there is a potential dangerous location, reduces the car speed according to the distance from that point. The safe speed v^{safe} is defined by the following equations:

$$v^{\text{safe}} = \frac{d^2}{k_1} + k_2 \tag{3}$$

$$k_1 = \frac{d^2_{\text{limit}}}{v^{\text{des}} - v_{\min}} \tag{4}$$

$$k_2 = v_{\min} \tag{5}$$

where: d is the distance between the vehicle and path location with bad surface conditions (dimension: (km)); d_{limit} is the limiting distance (dimension: (km)) from which the evaluation of safe speed starts; k_1 (dimension: (h)) and k_2 (dimension: (km/h)) depend on the desired speed v_{des} (dimension: (km/h)) at the limiting distance and on the minimum speed $v_{\min}$ (dimension: (km/h)) near the dangerous location.

PERFORMANCE EVALUATION

We first present an analytical performance evaluation framework of the DGT-based proactive neighbor localization protocol illustrated in Section 'Distributed Geographic Table'. The analytical results are confirmed by simulations, showing that the proposed framework is an effective and efficient approach to evaluate the number of discovery steps for highly precise (D4V-instrumental) neighbor localization. We then investigate, through simulations, a D4V-based application which allows vehicles to adapt their routes according to traffic information gathered from other vehicles in the area.

DGT-based proactive localization

Assuming that N peers are distributed within a square surface with side of length L (dimension: (km)), the corresponding peer spatial density, denoted as ρ, is N/L^2

(dimension: (vehicles/km^2)). If peers are static and uniformly distributed over the square surface, ρ is also the local peer spatial density. In the presence of node mobility (but still under the assumption that there are N peers in the square region of interest), the peer distribution is likely to be non-uniform: the corresponding peer spatial density can be heuristically estimated as $\gamma N/L^2$, where $\gamma \in \mathbb{R}^+$ is a compensation factor to take into account the fact that peers could be locally denser ($\gamma > 1$) or sparser ($\gamma < 1$) than the average value ρ.

Assume that, at a specific time, a peer wants to identify the available geographic neighbors within a circular region of interest. Such a region, centered at the peer, is denoted as $\mathcal{R}$ and its area is A. In general, within the region of interest of a peer there are two classes of neighbors: detectable (i.e., peers which can be detected by one or more other peers) and non-detectable (i.e., peers which cannot be detected by any other peer).

Assuming that peers are distributed according to a two-dimensional Poisson distribution[2] with parameter ρ, the average number of peers in the region $\mathcal{R}$ is $\overline{N}_{\text{tot}}^{(\mathcal{R})} = \rho \cdot A$. Let us denote by $x \in (0, 1)$ the percentage of non-detectable peers in the region $\mathcal{R}$ (i.e., there are, on average, $x \cdot \overline{N}_{\text{tot}}^{(\mathcal{R})}$ non-detectable peers). Assuming further that the number of detectable peers in R has a Poisson distribution with parameter $\rho \cdot A \cdot (1 - x)$, it follows that their average value is $\overline{N}_D^{(\mathcal{R})} = \rho \cdot A \cdot (1 - x)$.

As described in Section 'Distributed Geographic Table', during each step of the discovery procedure, a peer picks the closest α known neighbors (if available) and sends them simultaneous FIND_NODES requests centered in the peer's geographic location. The goal of the interrogating peer is to retrieve detectable peers in its region of interest. If, at the end of an iteration, no new peer is retrieved, the discovery process ends and will be rescheduled according to a specific strategy. In order to evaluate the number of discovered peers at each discovery iteration (without counting the same peer more than once), the α FIND_NODES requests, scheduled at each discovery step, must be taken into account considering not only the intersections between pairs of peers, but also the possible intersections between η-tuples ($2 \leq \eta \leq \alpha$) of peers originating from the α contacted peers. The total number of such intersections is

$$\sum_{i=1}^{\alpha} \binom{\alpha}{i} = 2^\alpha - 1. \tag{6}$$

In Fig. 3, an illustrative scenario with $\alpha = 3$ overlapping circular regions (associated to 3 contacted peers) is shown.

Since the intersection of α circular regions can be highly varying (depending on their relative positions), we simplify the analysis assuming that adjacent contacted peers are spaced by an angle $2\pi/\alpha$ and are positioned in the center of the corresponding radius of the circular region of interest of the analyzed peer that is performing discovery requests. We denote as A_j the sum of the areas of the intersection region shared only by the requesting peer and j contacted peers. In Fig. 3, the areas $\{A_1, A_2, A_3\}$ are indicated. Such areas can be computed using the Matlab library available at (*Vakulenko*). Explicit expressions

[2] This is an approximation. In fact, owing to node mobility, the local distribution is likely to be non-Poisson. However, as we will consider only average values, the Poisson approximation will be shown to be accurate.
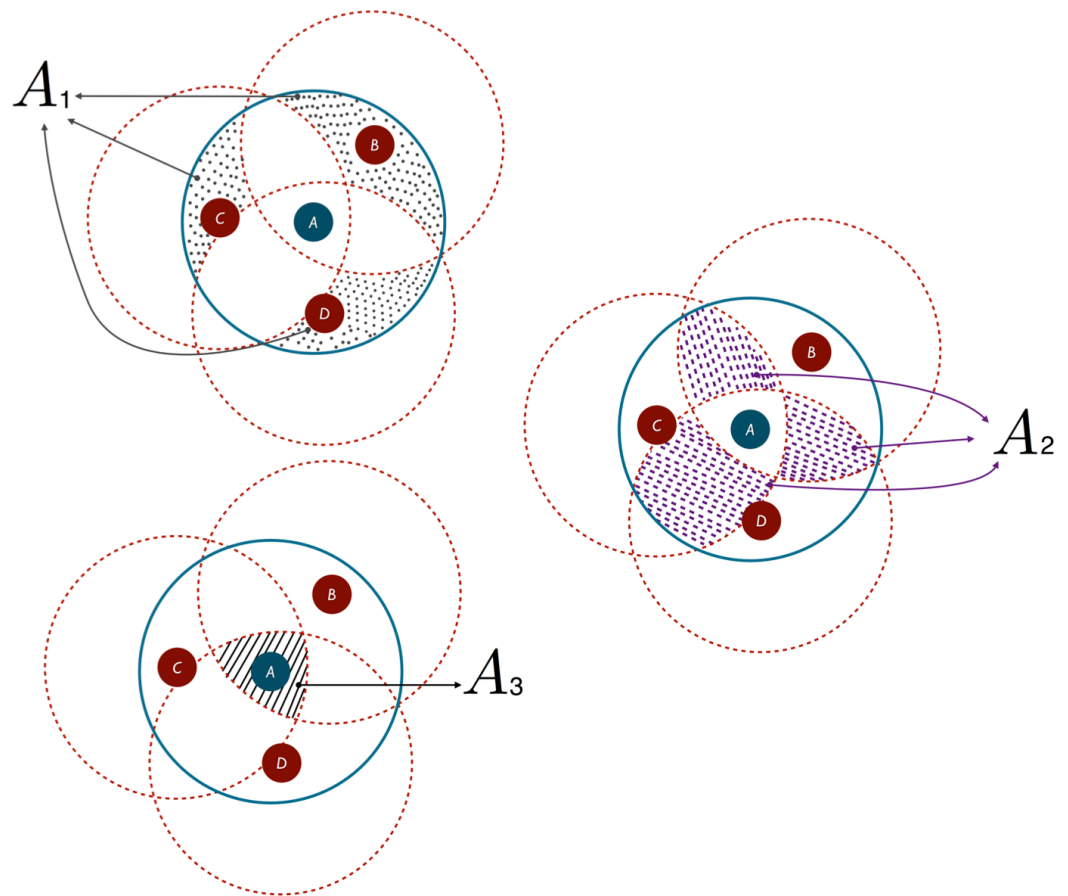
Figure 3 Intersection regions between 3 overlapping circular area of interest.

(not shown here for the sake of conciseness) can be derived according to the analysis in *Fewell (2006)*.

Under the above assumptions, the average number of new peers discovered after s steps can be written as

$$\overline{n}(s) = \begin{cases} l_0 & s = 1 \\ \overline{n}(s-1) + \sum_{j=1}^{\alpha} \overline{d}_j(\overline{n}(s-1)) & s > 1 \end{cases} \tag{7}$$

where: l_0 is the initial size of the peer list; $\overline{n}(1)$ is the average number of initial peers (transferred to the peer of interest); $\overline{n}(s-1)$ is the number of new peers discovered up to the $(s-1)$th step ($s \geq 2$); $\overline{d}_j(\overline{n}(s-1))$ represents the average number of new peers discovered in the region of area A_j upon querying $\overline{n}(s-1)$ peers and can be expressed as follows:

$$\overline{d}_j(\overline{n}(s-1)) = \rho \cdot A_j \cdot (1-x) \cdot b_j(\overline{n}(s-1)) \tag{8}$$

where $b_j(\overline{n}(s-1))$ is the probability that no replicas are obtained in the jth intersection between the applicant's region of interest and the regions of interest of the $\overline{n}(s-1)$ queried
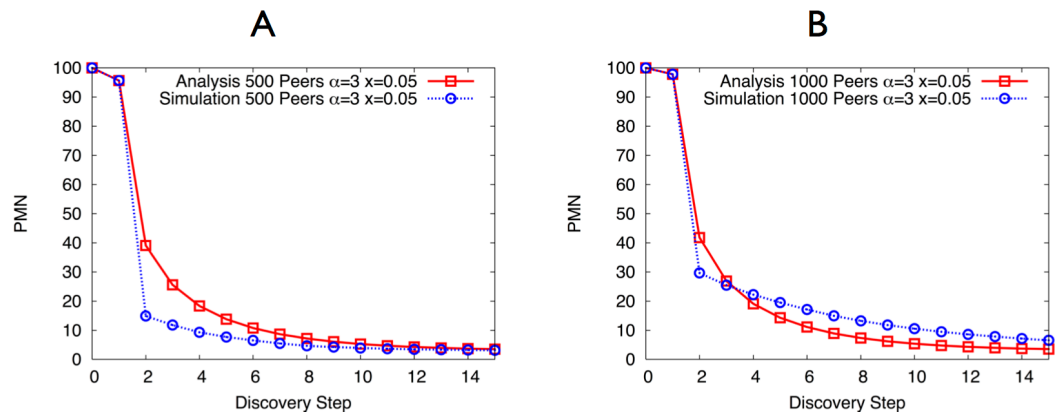
Figure 4 PMN as a function of the discovery step, considering (A) 500 active peers and (B) 1,000 active peers.

peers. This probability depends on: (i) the number of peers that share the same zone (i.e., j) and can answer with the same peer references; and (ii) the average number $\bar{n}(s-1)$ of known peers at step $s-1$—in fact, the number of known peers at each step needs to be taken into account to evaluate potential replicas. Considering the fact that if a peer knows its neighbors, the probability of discovering an already known peer is higher, the following heuristic expression for b_j will be shown to allow to derive accurate performance results:

$$b_j(\bar{n}(s-1)) = \left[1 - \frac{\bar{n}(s-1)}{\overline{N}_D^{(\mathcal{R})}}\right]^j. \tag{9}$$

Since j is the number of peers that share the same zone, in (9), by using j as exponent of the probability of receiving replicas, the higher is j, the lower is b_j.

Finally, the average number of newly discovered peers up to step s can be expressed as follows:

$$\bar{n}(s) = \bar{n}(s-1) + \sum_{j=1}^{\alpha} \rho \cdot A_j \cdot (1-x) \cdot \left[1 - \frac{\bar{n}(s-1)}{\overline{N}_D^{(\mathcal{R})}}\right]^j. \tag{10}$$

Note that the recursive analytical computation of $\{\bar{n}(s)\}$ stops when a pre-set peer discovery limiting number is reached.

In Fig. 4, performance results predicted by the analytical model proposed above are compared with simulation results obtained with DEUS (described in more detail in Section 'D4V performance evaluation'), considering scenarios with (a) 500 peers and (b) 1,000 peers. In both cases (a) and (b), peers are distributed within a square surface with side of length $L = 6.53$ km, with an initial peer list size with $\bar{n}(1) = l_0 = 10$ peers, a limiting number of discovered peers equal to 100, and $x = 0.05$. It can be observed that analytical performance results are very close to simulation results, so that one can conclude that the accuracy of the analytical framework is satisfactory.
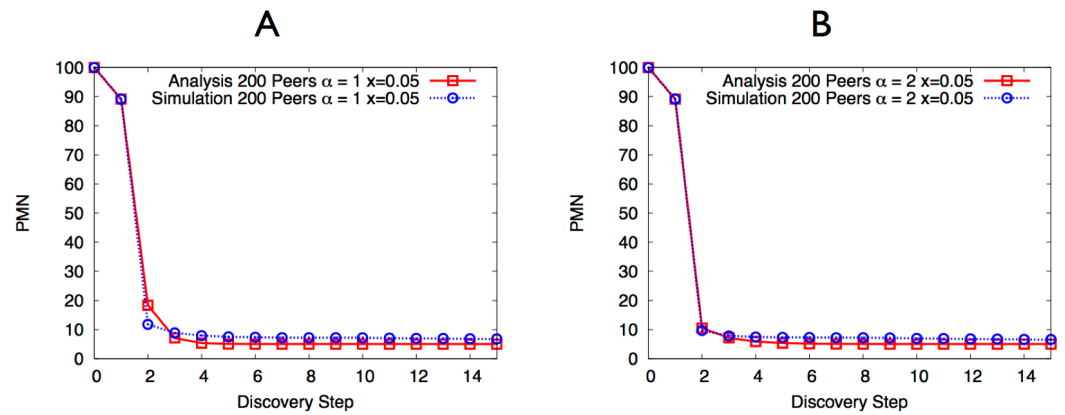
Figure 5 PMN as a function of the discovery step, considering (A) 1 and (B) 2 FIND_NODES requests at each neighborhood discovery step.

In order to investigate the impact of α, in Fig. 5 the Percentage of Missing Nodes (PMN) in the GBs of a peer, with respect to those actually present in the area, is shown, as a function of the discovery step, considering (a) $\alpha = 1$ and (b) $\alpha = 2$. In both cases, the number of active peers is set to 200. It can be observed that the agreement between simulation and analytical results is even higher than in Fig. 4. By observing the results in Figs. 5 and 4, it can be concluded that a small number of discovery steps (namely, 4) is sufficient, regardless of the value of α, to significantly reduce the PMN.

D4V performance evaluation

The performance evaluation of D4V has been mostly carried out by means of an extensive simulative analysis, complemented by a preliminary experimental evaluation of a D4V system prototype deployed on the PlanetLab global testbed. While a thorough performance evaluation of the proposed D4V system would require a wide range of (resource intensive) on-field experiments, we rely on the fact that discrete event simulations are deemed useful to provide a proof-of-concept in this domain (*Stojmenovic, 2008*). In particular, we focus on a 4G wireless communication scenario, based on the Long-Term Evolution (LTE) technology (*Cox, 2012*).

Simulation methodology

DEUS is an open source, Java-based, general-purpose discrete event simulation tool, which is particularly suitable for the application-level analysis of distributed systems with thousands of nodes, characterized by a high level of churn (node joins and departures) and reconfiguration of connections among nodes (*Amoretti, Agosti & Zanichelli, 2009*). On the other hand, ns-3 is a widely known open source tool for the discrete event simulation of Internet systems (focusing on low layers of the protocol stack, e.g., MAC and physical), which relies on high-quality contributions of the community to develop new models, to debug or to maintain existing ones, and to share results (*ns-3 Development Team*).

In *Amoretti et al. (2013)*, we describe a sound methodology to integrate DEUS (*Amoretti, Agosti & Zanichelli, 2009*) and ns-3 (*ns-3 Development Team*), leading

to a more accurate performance evaluation of large-scale mobile and distributed systems. The main steps of the co-simulation methodology proposed in *Amoretti et al. (2013)* can be summarized as follows:

1. given a complex system to be simulated, identify the main sub-system types, each one being characterized by specific networking parameters;
2. with ns-3: create detailed simulation models of the sub-systems (i.e., sub-models) and measure their characteristic transmission delays, taking into account both message payloads and proper headers;
3. with DEUS: simulate the whole distributed system, with refined scheduling of communication events, taking into account the transmission delays computed at step 2.

Regarding step 2, we have used ns-3's LENA LTE-EPC package,[3] by modifying the C++ class which creates the logs for the Radio Link Control (RLC) protocol. The modified class logs a discretized Probability Density Function (PDF) of the RLC packet delay. The latter is then used to generate realistic packet delays in the DEUS-based simulations, using the well-known *inversion method* (*Papoulis, 1991*). For practical implementation purposes, the discretized PDF of the downlink RLC packet delay is approximated by a piecewise constant function, whose numerical inversion is straightforward and computationally inexpensive.

We have simulated a D4V-based application deployed across the city of Parma, considering a number of vehicles that move over 100 km of realistic paths generated using the Google Maps API. Each simulated vehicle selects a different path and starts moving over it. Using the features provided by the Google Maps API, we have created a simple HTML & Javascript control page, which allows to monitor the time progression of the simulated system, where any peer can be selected to view its neighborhood: a few video demos are publicly available (*Distributed System Group*). The simulation covers 10 h of D4V system life (10000 virtual time units) with 20 SSs, 5 virtual paths with bad road surface (due to either ice, water, snow, or pothole), accident events characterized by a Poisson arrival process, and with different message types to disseminate information about sensed traffic data. Simulations with DEUS have been repeated with different seeds for the random number generator, to obtain a narrow I_{95} confidence interval. The performance results reported in Figs. 8–11 are obtained by averaging over the simulation runs, considering the whole set of simulated peers.

The considered simulation set-up is characterized by the DGT configuration which gives the best performance in urban scenarios (according to a previous study (*Picone, Amoretti & Zanichelli, 2011a*)), which is summarized in the following, using the DGT description formalism defined in Section 'Distributed Geographic Table'. Each peer has: $K = 4$ GBs, with the same thickness $r = 0.5$ km; a limiting number of discovered peers equal to 10; a region of interest of 12.5 km^2; an adaptive discovery period ranging from 1.5 min to 6 min, depending on the number of new discovered peers during each iteration step. The discovery period of a peer is an increasing function of the degree of knowledge of its neighborhood, corresponding to the decrement of the number of new discovered peers in the same area of interest.

[3] We used the version released the 23rd of January 2013 (*LTE-EPC Network Simulator*).
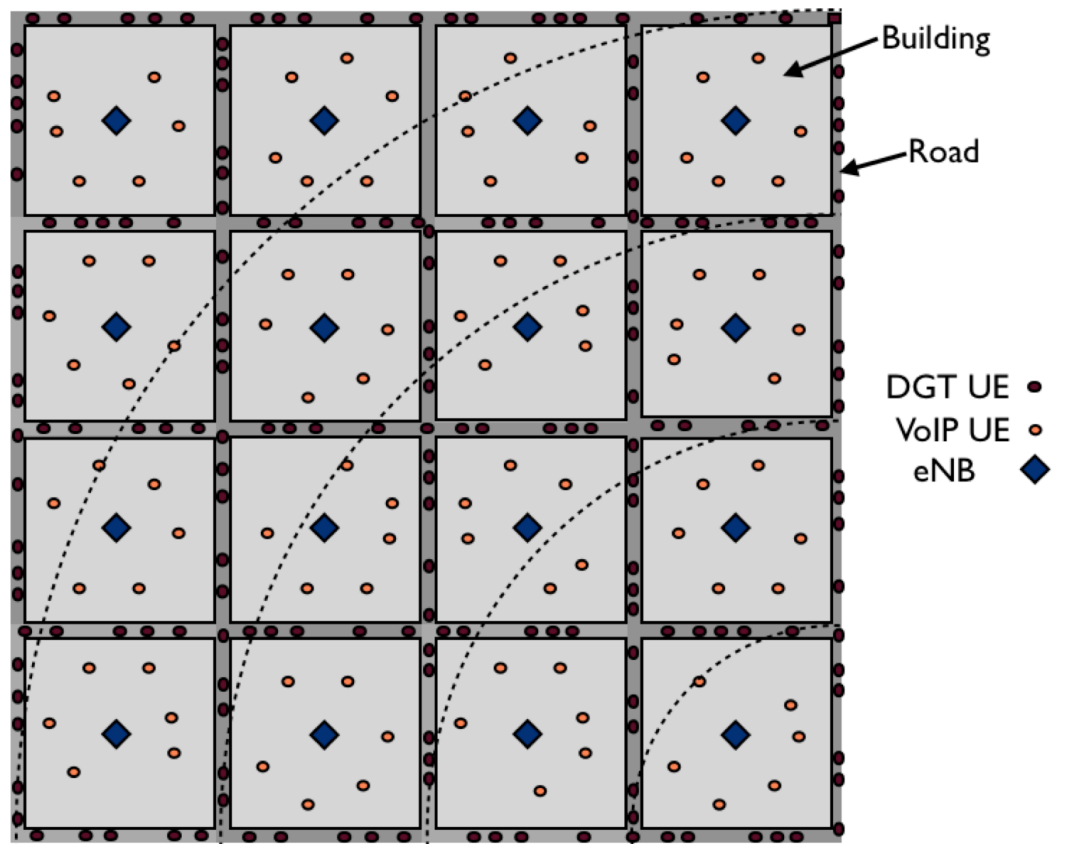
Figure 6 Bird's-eye view of the simulated scenario.

The transmission delay of a DGT packet has been computed by simulating with ns-3 the sub-system illustrated in Fig. 6 (averaging over 20 simulation runs). To match the previously described DGT configuration (i.e., DGT peers having GBs that cover a circular region of interest with radius equal to 2 km), we consider a square region having side length $l = 2$ km, with a grid of $n_r = 10$ roads (5 in the N-S direction, and 5 in the W-E direction) and vehicles running over them (with linear density δ). The total amount of DGT User Equipments (UEs) can be expressed as $n = n_r \delta l$. Parallel roads are spaced by $l/4 = 0.5$ km. In the map, there are 16 large buildings with square footprint, and seven floor-tall. Within each building, there are $n_v/16$ other randomly located UEs, where n_v is the total number of UEs in all buildings. The path loss model is ns3::BuildingsPropagationLossModel.

On top of each building, exactly in the middle, there is an Evolved NodeB (eNB), i.e., a base station which serves a subset of the $n + n_v$ UEs.[4] The configuration of the eNBs includes FDD paired spectrum, with 50 Resource Blocks (RBs) for the uplink (which corresponds to a nominal transmission rate of 50 Mbps) and the same for the downlink—this is coherent with currently deployed LTE systems. DGT UEs use UDP to send four types of DGT packets to each other. The first type, called Descriptor (24 bytes), is for neighborhood consistency maintenance purposes. The second type of packet, the Lookup Request (20 bytes), is used to search for remote peers placed around a specified

[4] Such a dense deployment of eNBs may appear to be quite optimistic, but it represents a realistic scenario for medium-term UE systems.
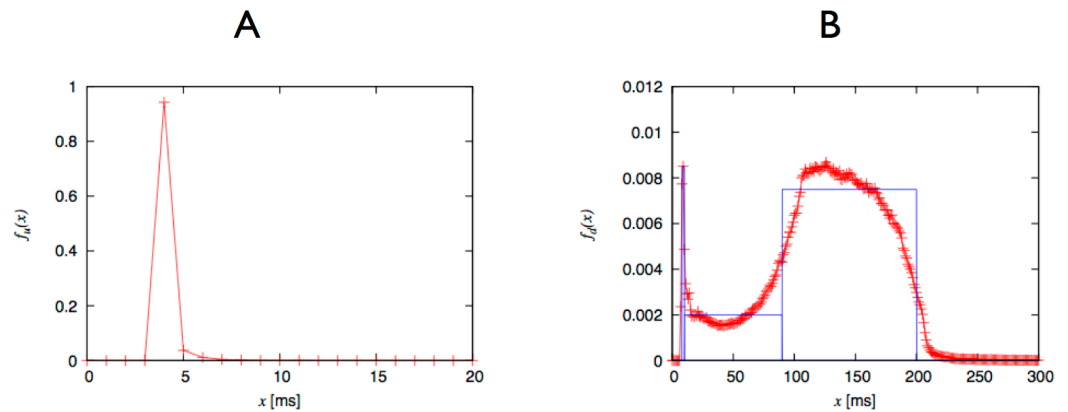
Figure 7 PDFs of the uplink (A) and downlink (B) delays for DGT packets.

location. The third packet type is the Lookup Response (500 bytes), which is sent by a DGT peer as a reply to a lookup request, if the peer owns the searched resource or information. Finally, the fourth type of packet is related to traffic information (66 bytes). All packet types have also a 12 byte header. We set an inter-packet interval of 50 ms for all types of DGT messages. Thus, the maximum and minimum rates are, respectively, $512 \times 20 \simeq 10$ kB/s and $32 \times 20 = 0.64$ kB/s. In a dynamic DGT scenario (the one simulated with DEUS), packets are not sent periodically—descriptors are sent only every ϵ meters; lookup requests (as well as lookup responses) are sent only when necessary; traffic information messages are sent only when something "interesting" can be communicated to the other peers (for example, a traffic jam or an incident). In order to simulate the presence of non-DGT traffic over LTE networks, we also include $n_v = 96$ other UEs, transmitting and receiving VoIP packets (using UDP) with a remote host located in the Internet. These packets have a 12 byte header and a 13 byte payload, with inter-packet interval set to 20 ms (the AMR 4.75 kbps codec is considered). The PDF of the resulting uplink delay, shown in Fig. 7A, can be approximated as a Dirac delta function. The PDF of the downlink delay, shown in Fig. 7B, can instead be approximated with a piecewise constant function, with three levels.

Such delay profiles scale from small scenarios to larger ones, as they refer to intra-GB communications only. A DGT message can be propagated across the whole city, from one peer to another, relayed by intermediate peers. Each message propagation would be affected only by the data traffic within the GB of the forwarding peer—where the obtained delay profiles apply.

Parameters for large-scale analysis

The following set of performance metrics has been considered in the DEUS-based simulations of the D4V system:

- CP (dimensionless): Estimated Coverage Percentage of D4V messages (*TrafficInformation* and *SensorData*) at a certain time of the simulation. It is evaluated as the ratio

Table 1 Parameters affecting the performance of a D4V system.

Symbol	Description	Values
K	Number of GBs	4
r	Thickness of each GB	0.5 km
ϵ	Position update threshold	[0.1; 1] km
ER	Event range	[1; 10] km
δ	Peer spatial density	[5; 20] veh/km
P	Packet loss percentage	[1; 10]

between the number of peers that actually received a specific message and the number of those which should have received it.

- PVTJ (dimensionless): Average percentage of vehicles (with respect to the total number of vehicles) involved in a traffic jam.

- DFE (dimension: (km)): the Distance From Event is the average distance from a traffic jam of interested vehicles which have not received the information about the traffic jam yet. The higher the DFE, the higher the security margin (and related time) to receive the message.

- DR (dimension: (kB/s/peer)): average data rate per peer.

Table 1 summarizes the values of the main parameters that affect the performance of the D4V system.

Impact of ϵ

The first step of the simulation-based D4V evaluation aims at analyzing the impact of the value of the threshold ϵ, considering two representative values for the peer spatial density δ (namely, 10 veh/km and 20 veh/km), an Event Range ER equal to 4 km, and a packet loss percentage $P = 1\%$. As defined in Section 'Distributed Geographic Table', ϵ represents the minimum displacement threshold considered by a peer to notify its geographic position update to the peers in its neighborhood. Our performance analysis aims at evaluating the effects of the variation of the update frequency on the information dissemination and, consequently, on the system performance. In Fig. 8, the impact of ϵ on the considered performance metrics is evaluated. In particular, Fig. 8A, where the CP is investigated, shows that traffic information messages are highly distributed to active peers in the region of interest. As expected, a higher peer spatial density contributes to increase knowledge sharing, thus increasing the CP. In Fig. 8B, the PVTJ is investigated, showing the inherent robustness of D4V. In fact, even in the presence of a reduced update frequency, D4V can properly distribute traffic information, leaving only a small percentage of drivers in traffic jams. We remark that even with lower peer spatial densities (such as $\delta = 4$ peers/km) the performance would not change, provided that ϵ is properly configured (as will be shown in Fig. 10B, discussed below). The effectiveness of the D4V approach is further shown by the DFE results in Fig. 8C. In particular, the DFE remains approximatively constant and very close to the dissemination range value of 4 km, confirming that peers that do not
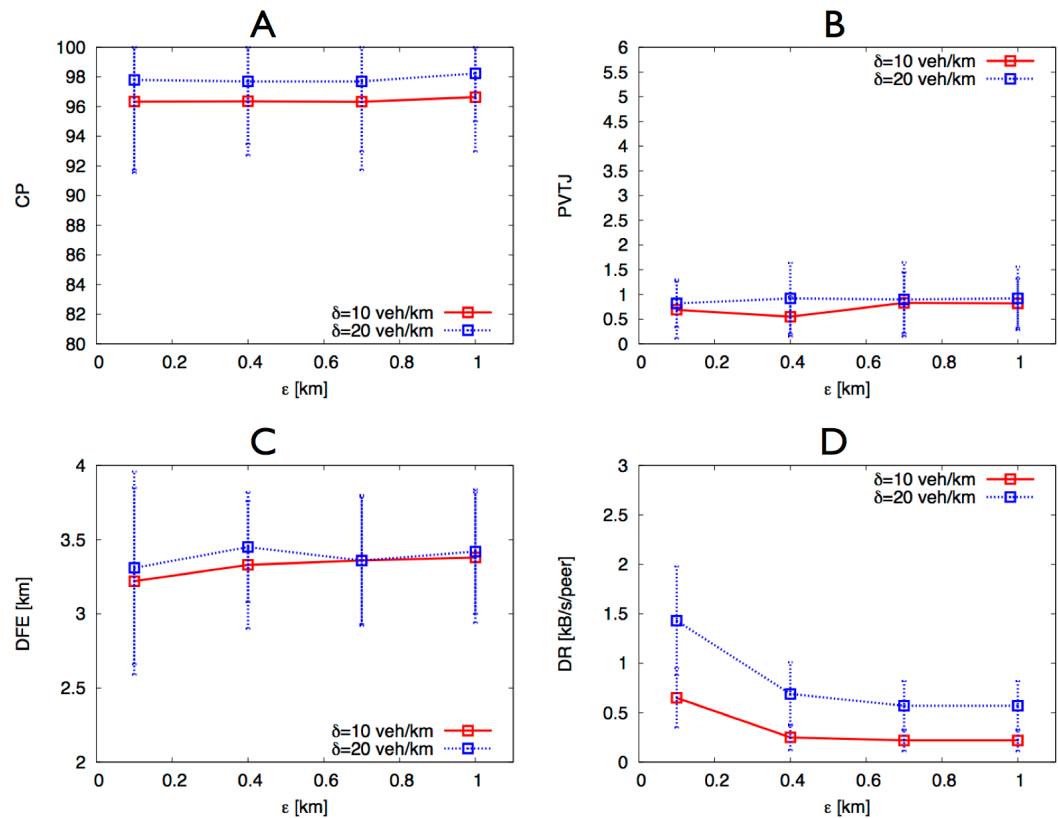
Figure 8 Simulation results for different values of the position update threshold.

receive a traffic message are those located very far from the traffic event, thus still having a high probability of receiving the alert on time. This analysis suggests that vehicles stuck in traffic jams are the ones really close to the traffic jam and with not enough time to react and change direction. From the results in Fig. 8D, it can be observed that a finer position update rate (i.e., a lower value of ϵ) and/or a higher peer spatial density increase DR.

Impact of event range

In Fig. 9, the same performance metrics of Fig. 8 are investigated as functions of ER, with $\epsilon = 1$ km and $P = 1\%$. The results in Figs. 9A and 9B show clearly that a short ER (as small as 1 km) affects the message distribution process, due to a lower margin between the traffic jam and the drivers. In this situation, peers may receive alert messages when they are too close to dangerous situations, thus becoming involved in the queue. In particular, we remark how a lower vehicle density worsens such a phenomenon due to the smaller number of peers available in their knowledge database which can redistribute traffic condition event messages. At the same time, it can be observed how there is no significant gap using ER values larger than 4 km for both peer density curves. In Fig. 9C, the DFE value for the considered configurations is shown as a function of the range. For comparison, the optimal distance from the event is also shown. The latter coincides with the value of the Event Range as, ideally, the minimum distance of peers which did not receive the traffic information message yet is clearly the range of interest. The obtained results show that
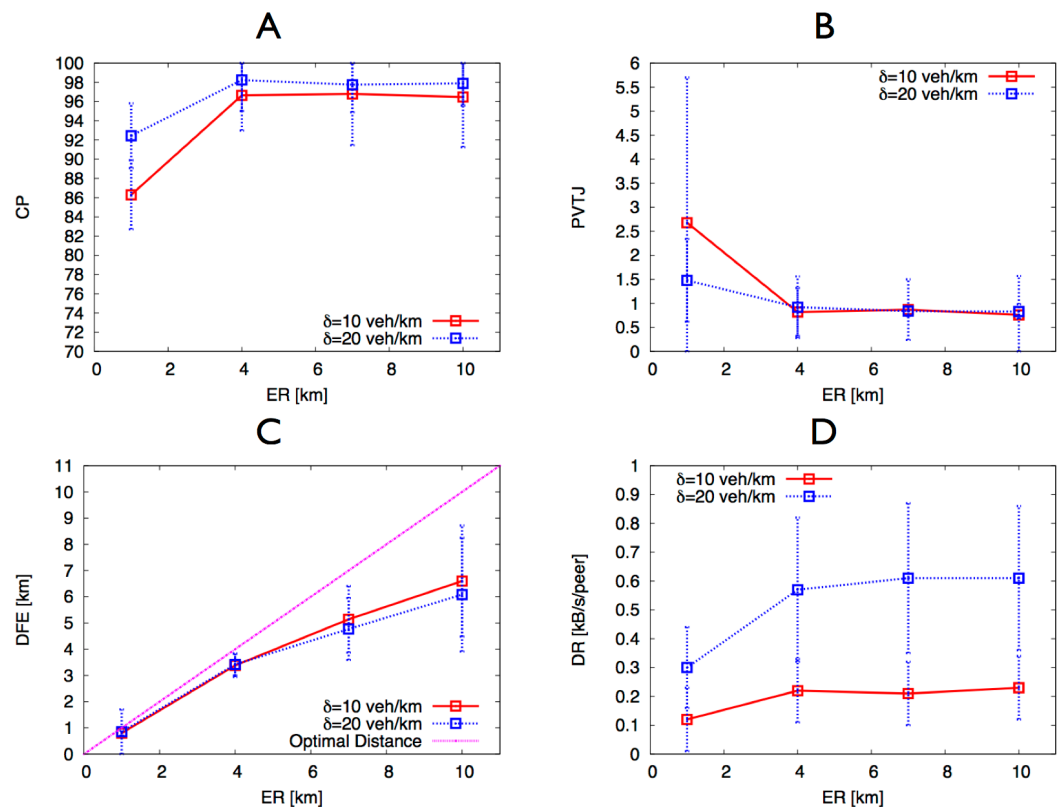
Figure 9 Results for different values of the dissemination range.

within a 4 km Event Range, the DFE remains very close to the optimal bound, while it increasingly separates from it for higher ER values. This is quite reasonable, as the area to be covered increases with the square of the Event Range. Hence, drivers who do not receive the alert for a specific event are in any case sufficiently far from it and will receive the alert with enough time to react. Finally, an extended ER corresponds to an increased notification area and, consequently, to a larger number of interested drivers that may be contacted. However, as shown in Fig. 9D, this slightly affects the amount of exchanged messages.

Impact of peer spatial density

The third stage of the simulative analysis aims at evaluating the impact, on system performance, of the peer spatial density. In all considered cases, $\epsilon = 1$ km, ER = 4 km, and $P = 1\%$. The scenario is characterized by an initially growing number of active vehicles, followed by a stable phase without new joins or disconnections. The results in Fig. 10A confirm that the proposed solution copes with different peer spatial densities with no performance degradation, keeping the CP significantly high (between 98% and 100%) even in the case of very low density (5 peers/km), which could be quite critical for VANET-based applications. We recall that, if a mobile peer finds itself in a desert region, it will still be able to fill its external GB with remote peers, by requesting their contacts to the bootstrap node (described in Section 'Distributed Geographic Table'), as if it is joining again the network with a new geographic location. Such a distributed knowledge
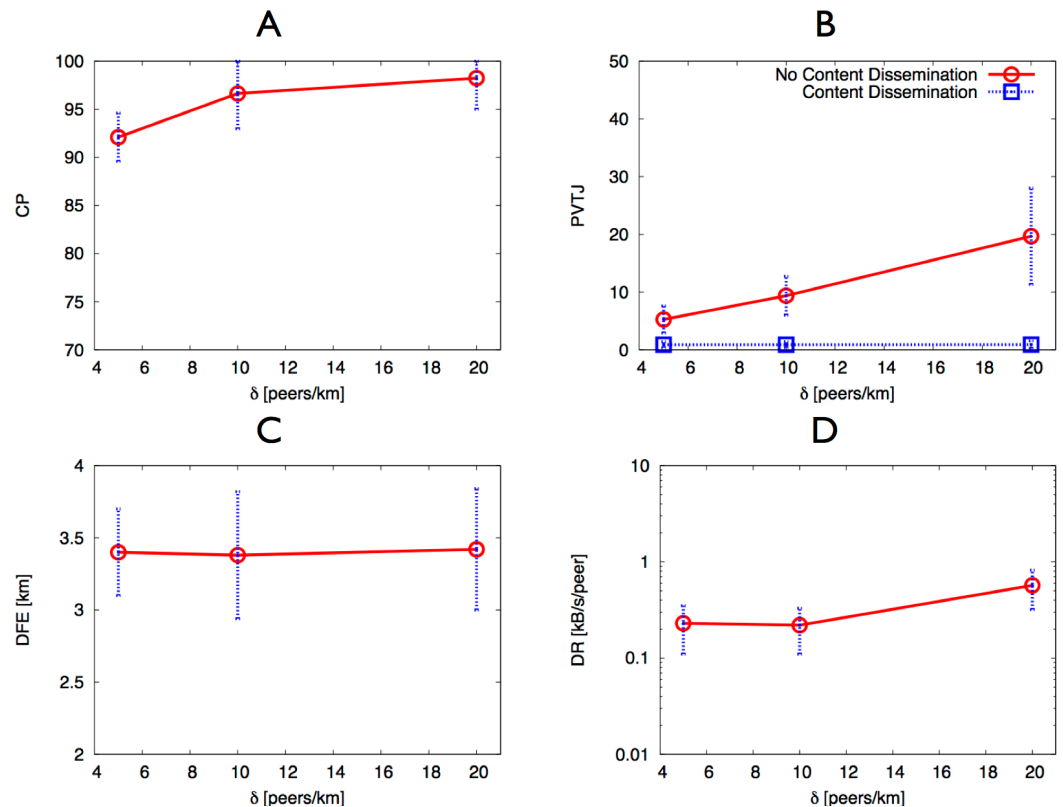
Figure 10 Simulation results for different peer densities.

provides appropriate support to efficiently disseminate messages about traffic jams or sensed data. As already observed in Section 'Impact of event range', the results in Fig. 10C show that an increasing number of active peers maintains the DFE high and close to the dissemination range. This results in an accurate dissemination of traffic information messages that allows drivers to receive alert information on time, still sufficiently far from the dangerous location.

In Fig. 10B, the percentages of vehicles blocked in a traffic jam, with and without D4V content dissemination, are directly compared. These results confirm that the D4V approach drastically reduces the number of involved vehicles that would otherwise grow significantly for increasing density.

In Fig. 10D, the average data traffic per peer (dimension: (kB/s/peer)), required to maintain the DGT overlay and disseminate traffic information messages to other active neighbors, is shown as a function of the peer spatial density. Since UDP is the used transport protocol, there are no retransmissions in the presence of lost packets—more details can be found in *Picone, Amoretti & Zanichelli (2011b)*. Here, we investigate the average bandwidth—estimated from simulative results, corrected considering the cost of headers—consumed in the best case (when the transmitted message is much longer than the IP header) and in the worst case (when the transmitted message has a size comparable to that of the IP header, e.g., a location update, which contains only a peer descriptor and

a location). Even if there is an unavoidable growth for increasing peer spatial densities, the amount of data exchanged by each peer remains limited. This behavior is associated with the fact, described in Section "Impact of event range", that the D4V system uses an opportunistic content dissemination strategy. In fact, this approach tries to minimize the amount of transmitted packets, by forwarding them only to interested users, trying, at the same time, to reduce the number of duplicated messages. The considered values of peer spatial density are 5 veh/km, 10 veh/km, and 20 veh/km. Higher values would be neither realistic nor interesting, as they would mean that all vehicles on the roads run the DGT.

A similar analysis has been carried out by *Heep et al. (2013)*, regarding Overdrive—the most recent location-aware P2P overlay scheme for smart traffic applications (as anticipated in Section 'Related work'). Their Geographic Unicast Message Success Rate (GUMSR) can be compared to our CP. Overdrive is characterized by GUMSR between 90% and 95%, with DR > 1 kB/s/peer. The D4V shows a CP > 95% with DR < 1 (kB/s/peer). In Overdrive, the bandwidth consumption depends on the flooding rate. In D4V, message dissemination is mostly affected by peer spatial density (for which GBs could be more or less filled) and by the dissemination range. As shown in Section 'Impact of event range', even large variations of the latter parameter keep DR significantly below 1 (kB/s). Moreover, in our analysis, we investigate the DFE of the 100-CP % of the peers that do not receive the notifications. The higher the DFE, the higher the distance between the event's location and the peers that have not been notified, the higher the security margin. To summarize, in order to have a comprehensive behavior and performance evaluation of location-aware P2P overlay schemes, CP and DFE must be jointly investigated.

Robustness

In Fig. 11, the robustness is investigated by analyzing the impact of the packet loss percentage P on CP, PVTJ, DFE, and DR. In all cases, $\epsilon = 1$ km and ER $= 4$ km. In the current simulator, there is no recovery procedure to verify whether a transmitted message has been correctly delivered and, if necessary, to retransmit it. This needs to be taken into account to properly interpret the obtained results, in particular for the dissemination of traffic information messages and the global robustness of the DGT approach.

In Fig. 11A, the global CP is shown as a function of the packet loss percentage, confirming that peers maintain a detailed knowledge of traffic events (on average more than 90%) in the first GB. In Fig. 11B, the PVTJ appears as a slightly increasing function of P, given that some peers may not receive alerts on the dangerous event and could be stuck in a queue. The distributed knowledge provided and maintained by the DGT allows to inform a large number of drivers, thus keeping the number of queued vehicles really small. The robustness of D4V is also confirmed by results in Fig. 11C, showing that the peers that do not receive traffic information messages related to a dangerous event are considerably distant from the event's location. Moreover, the DFE is almost independent of P. Finally, in Fig. 11D the DR is shown as a function of P. It can be observed that the DR is unavoidably lower than in the other scenarios, due to the lack of a recovery procedure for lost packets.
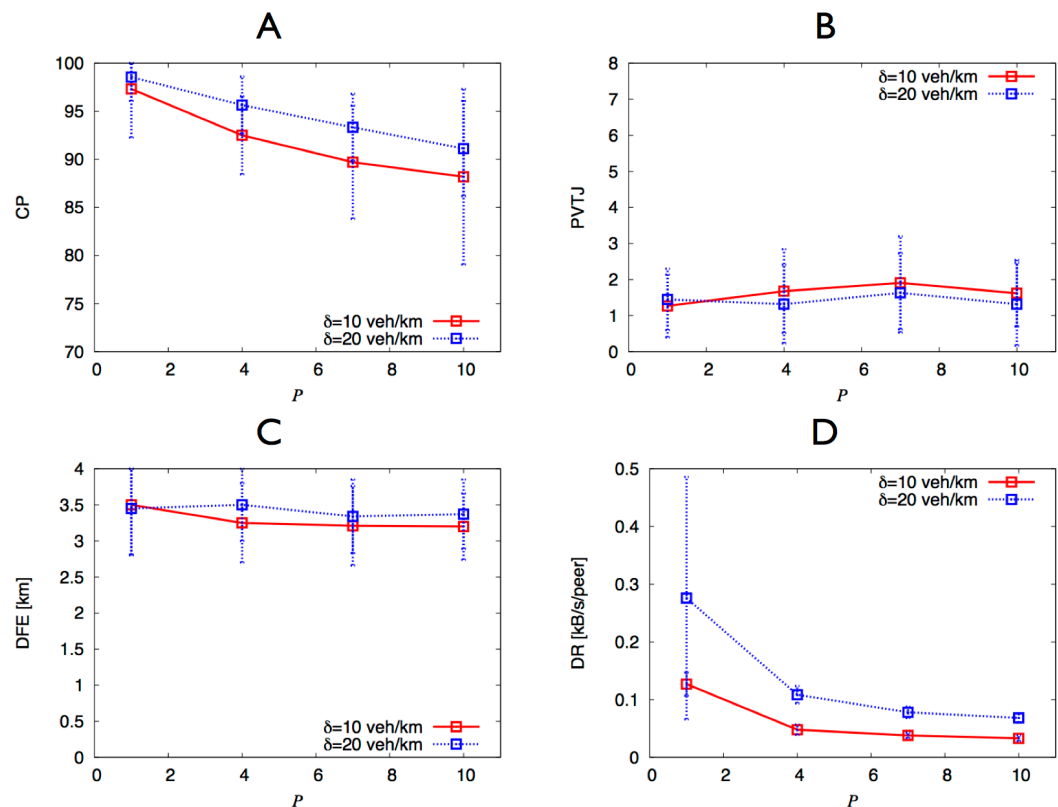
Figure 11 Simulation results for different packet loss percentages.

Vehicle speed analysis

Finally, considering the behavioral model of a driver in proximity of a road stretch with a bad surface condition, that we have introduced in Section 'Mobility Model', in Fig. 12 we show the monitored speed for five virtual tracks with bad surface conditions for all drivers (including both the informed ones and those not informed). The observed results clearly show that a decreased speed is measured near the critical location (at distance zero), along with an increasing speed while moving away from it. Owing to this behavior, it can be concluded that the deployment of D4V would probably reduce the risk of accidents and nuisances, on account of the D4V-based information sharing among drivers, especially those approaching the dangerous event.

Experimental evaluation

The extensive simulative analysis of the DGT gave us valuable feedback for the development of a DGT Java library and a first prototype of the D4V traffic information system. The DGT library implements the core functionalities such as neighborhood and management, as well as GB maintenance. A D4V application layer uses such features to implement the content dissemination algorithm and the user interface to collect inputs related to a specific traffic event, and to show approaching dangerous situations. The development of the DGT library is based on the open source peer-to-peer middleware
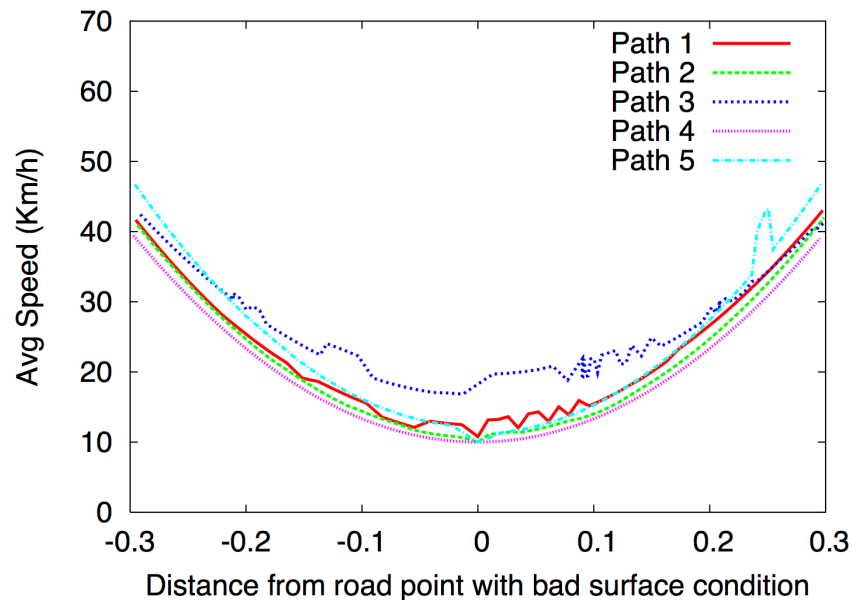
Figure 12 Average of driver speed near road points with bad surface condition.

called *Sip2Peer* (*Sip2Peer*), which provides SIP-based primitives for the implementation of any peer-to-peer overlay scheme and application.

In order to properly measure the network performance, to understand if the results of the simulation analysis are confirmed in a real distributed environment, we deployed D4V nodes on PlanetLab,[5] which is a global research network that supports the development of new network services. PlanetLab currently consists of about 1,089 nodes at 532 sites (the University of Parma contributes with 2 nodes).

In detail, we deployed 50 D4V peers on 13 different PlanetLab servers, located in 13 different countries. Every 30 s, each node logs all the required information (e.g., geographic location, exchanged kbytes, received and sent messages) to analyze the behavior of the node. At the end of each experiment, a dedicated tool parses all available log files, to build a time line of the experiment made by steps of 30 s containing all the required statistics for the performance evaluation. All experiments have been run several times.

Figure 13A illustrates the Coverage Percentage. The generation of traffic messages starts 400 s after the activation of a dedicated D4V event-generator node, in order to give them sufficient time to build the DGT overlay. Results show that the average value of CP is very high (close to 97%), and in particular significantly near the average value of our simulations (≃98%). The CP curve shows that when new messages are generated, the coverage percentage goes lightly down to lower values but after one or two time line steps (30/60 s) recovers to a high coverage percentage, thus confirming that the dissemination process and the neighborhood knowledge allow to efficiently distribute messages.

Figure 13B shows the DFE of the PlanetLab deployment, considering a 4 km interest range ffor disseminated messages. The graph confirms that vehicles that did not receive the message are on average significantly far from the dangerous event, and with an

[5] https://www.planet-lab.org/
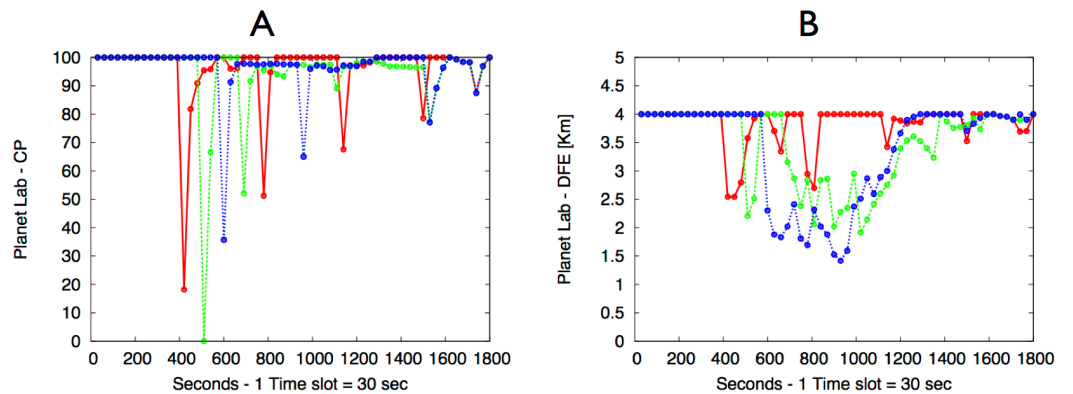
Figure 13 PlanetLab experimental results: (A) Coverage Percentage; (B) Distance From Event (DFE).

high probability have a sufficient margin to receive the message before approaching the potentially dangerous location, by changing their direction to reach their destination using a different route, or just adapting their vehicle speed (for example, in proximity of a portion of damaged road surface).

CONCLUSIONS

In this paper, we have introduced D4V, a scalable system for opportunistic dissemination of information gathered through commercial smartphones, from vehicle sensors and driver inputs. D4V relies on the potential of DGT, a P2P overlay network which unifies the concepts of geographical and virtual neighborhoods.

Two key results have been presented. The first one is given by the derivation of an analytical framework to characterize the discovery procedure of the DGT proactive neighbor localization protocol. The outcome, namely the average number of newly discovered nodes at each step, can provide useful guidelines for the design of a DGT-based application to determine how to appropriately set the main system parameters in order to guarantee a desired missing node percentage. The second result is given by the design of an effective and efficient opportunistic dissemination strategy which relies on the DGT to distribute vehicular information and sensed data to interested drivers.

Simulation results show that the proposed D4V system guarantees a high vehicular notification coverage, over a wide range of system parameters' values, whilst generating limited control data traffic and coping reasonably well with significant packet losses. Hence, we are confident that D4V could be effectively used on the road to reduce the number of drivers involved in traffic jams, as well as to disseminate alert messages about potentially dangerous road stretches, thus allowing drivers to reduce risks and nuisances along their paths.

Further work will investigate the optimization of opportunistic message dissemination at the minimum D4V message traffic load (e.g., by estimating vehicle trajectories). Moreover, we will investigate a global communication model which takes into account

both user mobility and available wireless network (Wi-Fi and cellular) coverage: this will likely improve the flexibility, accuracy, and reliability of D4V.

Funding

The work of G Ferrari was partially supported under the one-year project "Cross-Network Effective Traffic Alerts Dissemination" (X-NETAD, Eureka Label E! 6252, 2011–2012), 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." The work of Marco Picone is supported by Guglielmo srl, Reggio Emilia, Italy. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Ministry of Foreign Affairs.
The Israeli Industry Center for R&D (Israel).
Guglielmo srl, Reggio Emilia, Italy.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Marco Picone conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, performed the computation work.
- Michele Amoretti conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, performed the computation work.
- Gianluigi Ferrari and Francesco Zanichelli conceived and designed the experiments, reviewed drafts of the paper.

Data Availability

The following information was supplied regarding the deposition of related data:
https://github.com/dsg-unipr/deus/.

REFERENCES

Aberer K, Alima LO, Ghodsi A, Girdzijauskas S, Haridi S, Hauswirth M. 2005. The essence of P2P: a reference architecture for overlay networks. In: *5th IEEE Int.'l conference on peer-to-peer computing (P2P 2005)*. Konstanz, Germany.

Amoretti M. 2009. A survey of peer-to-peer overlay schemes: effectiveness, efficiency and security. *Recent Patents on Computer Science* **2(3)**:195–213 DOI 10.2174/2213275910902030195.

Amoretti M, Agosti M, Zanichelli F. 2009. DEUS: a discrete event universal simulator. In: *2nd ICST/ACM Int.'l conference on simulation tools and techniques (SIMUTools 2009)*. Roma, Italy.

Amoretti M, Picone M, Zanichelli F, Ferrari G. 2013. Simulating mobile and distributed systems with DEUS and ns-3. In: *International conference on high performance computing and simulation 2013*. Helsinki, Finland.

Bronsted J, Kristensen LM. 2006. Specification and performance evaluation of two zone dissemination protocols for vehicular ad-hoc networks. In: *39th annual simulation symposium, Huntsville*. Alabama, USA.

Buchanan M. 2003. *Nexus: small worlds and the groundbreaking theory of networks.* New York: W. W. Norton & Company.

Burke J, Estrin D, Hansen M, Parker A, Ramanathan N, Reddy S, Srivastava MB. 2006. Participatory sensing. In: *WSW 2006*. Boulder, Colorado, USA.

Caliskan M, Graupner D, Mauve M. 2006. Decentralized discovery of free parking places. In: *3rd Int.'l workshop on vehicular ad hoc networks*. Los Angeles, California, USA.

Cox C. 2012. *An introduction to LTE: LTE, LTE-advanced, SAE and 4G mobile communications.* Hoboken: Wiley.

Dietzel S, Petit J, Kargl F, Scheuermann B. 2014. In-network aggregation for vehicular ad hoc networks. *IEEE Communications Surveys & Tutorials* **16(4)**:1909–1932 DOI 10.1109/COMST.2014.2320091.

Distributed System Group. D4V Videos. *Available at http://dsg.ce.unipr.it/d4v*.

Du L, Dao H. 2015. Information dissemination delay in vehicle-to-vehicle communication networks in a traffic stream. *IEEE Transactions on Intelligent Transportation Systems* **16(1)**:66–80 DOI 10.1109/TITS.2014.2326331.

eMule project. Homepage. *Available at http://www.emule-project.net*.

Fewell MP. 2006. Area of common overlap of three circles. Tech. Report of the Department of Defence, Australian Government.

Fiore M, Harri J, Filali F, Bonnet F. 2009. Vehicular mobility simulation with VanetMobiSim. *Simulation* **87(4)**:275–300 DOI 10.1177/0037549709345997.

Fujiki T, Kirimura M, Umedu T, Higashino T. 2007. Efficient acquisition of local traffic information using inter-vehicle communication with queries. In: *10th Int.'l IEEE conference on intelligent transportation systems (ITCS '07)*. Seattle, Washington, USA.

Gerla M, Kleinrock L. 2011. Vehicular networks and the future of the mobile internet. *Computer Networks* **55(2)**:457–469 DOI 10.1016/j.comnet.2010.10.015.

Granovetter M. 1973. The strength of weak ties. *American Journal of Sociology* **78(6)**:1360–1380 DOI 10.1086/225469.

Hadaller D, Keshav S, Brecht T, Agarwal S. 2007. Vehicular opportunistic communication under the microscope. In: *Int.'l conference on mobile systems, applications and services (MobiSys)*. San Juan, Puerto Rico.

Han M, Moon S, Lee Y, Jang K, Lee D. 2008. Evaluation of VoIP quality over WiBro. In: *Passive and active measurement conference (PAM)*. Cleveland, Ohio, USA.

Harri J, Filali F, Bonnet C. 2005. A framework for mobility models generation and its application to inter-vehicular networks. In: *3rd IEEE Int.'l workshop on mobility management and wireless access*. Cologne, Germany.

Hartenstein H, Laberteaux K. 2009. *VANET vehicular applications and inter-networking technologies.* Hoboken: Wiley.

Heep B, Florian M, Volz J, Baumgart I. 2013. OverDrive: an overlay-based geocast service for smart traffic applications. In: *10th annual conference on wireless on-demand network systems and services (WONS)*. Banff, Alberta, Canada.

Hull B, Bychkovsky V, Zhang Y, Chen K, Goraczko M, Miu A, Shih E, Balakrishnan H, Madden S. 2006. Cartel: a distributed mobile sensor computing system. In: *4th ACM conference on embedded networked sensor systems (SenSys)*. Boulder, Colorado, USA.

Jedrzej R, Scheuermann B, Koegel M, Mauve M. 2009. PeerTIS: a peer-to-peer traffic information system. In: *Int.'l conference on mobile computing and networking*. Beijing, China.

Jiang D, Taliwal V, Meier A, Holfelder W, Herrtwich R. 2006. Design of 5.9 GHz DSRC-based vehicular safety communication. *IEEE Wireless Communications* **13(5)**:36–43 DOI 10.1109/WC-M.2006.250356.

Krauss S, Wagner P, Gawron C. 1997. Metastable states in a microscopic model of traffic flow. *Physical Review E* **55**:55–97 DOI 10.1103/PhysRevE.55.5597.

Lee U, Gerla M. 2010. A survey of urban vehicular sensing platforms. *Computer Networks* **54(4)**:527–544 DOI 10.1016/j.comnet.2009.07.011.

Lee U, Magistretti E, Zhou B, Gerla M, Bellavista P, Corradi A. 2006. MobEyes: smart mobs for urban monitoring with vehicular sensor networks. *IEEE Wireless Communications* **13(5)**:52–57 DOI 10.1109/WC-M.2006.250358.

Leontiadis I, Mascolo C. 2007. Opportunistic spatio-temporal dissemination system for vehicular networks. In: *Int.'l conference on mobile systems, applications and services (MobiSys)*. San Juan, Puerto Rico.

Longley PA, Goodchild MF, Maguire DJ, Rhind DW. 2005. *Geographic information systems and science*. Hoboken: Wiley.

LTE-EPC Network Simulator (LENA). release M5. *Available at http://mailman.isi.edu/pipermail/ ns-developers/2013-January/010819.html*.

Mahmoud A, Olariu S. 2007. Zipper: a zero-infrastructure peer-to-peer system for VANET. In: *Int.'l workshop on modeling analysis and simulation of wireless and mobile systems (MSWiM)*. Chania, Crete Island, Greece.

Maymounkov P, Mazieres D. 2002. Kademlia: a peer-to-peer information system based on the XOR metric. In: *1st international workshop on peer-to-peer systems*.

Meneguette RI, Maia G, Madeira ERM, Loureiro AAF. 2014. Autonomic data dissemination in highway Vehicular Ad Hoc Networks with diverse traffic conditions. In: *IEEE symposium on computers and communication (ISCC)*. Madeira, Portugal.

Mohan P, Padmanabhan V, Ramjee R. 2008. Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In: *4th ACM conference on embedded networked sensor systems (SenSys)*. Raleigh, North Carolina, USA.

Mun M, Reddy S, Shilton K, Yau N, Boda P, Burke J, Estrin D, Hansen M, Howard E, West R. 2009. PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. In: *Int.'l conference on mobile systems, applications and services (MobiSys)*. Krakow, Poland.

ns-3 Development Team. ns-3 Official Homepage. *Available at http://www.nsnam.org*.

Papoulis A. 1991. *Probability, random variables, and stochastic processes*. Milano: McGraw Hill.

Picone M, Amoretti M, Zanichelli F. 2011a. Proactive neighbor localization based on distributed geographic table. *International Journal of Pervasive Computing and Communications* **7(3)**:240–263 DOI 10.1108/17427371111173022.

Picone M, Amoretti M, Zanichelli F. 2011b. Evaluating the robustness of the DGT approach

for smartphone-based vehicular networks. In: *IEEE workshop on user mobility and vehicular networks*. Bonn, Germany.

Qureshi A, Carlisle J, Guttag J. 2006. Tavarua: video streaming with WWAN striping. In: *ACM multimedia*. ACM.

Rybicki J, Scheuermann B, Kiess W, Lochert C, Fallahi P, Mauve M. 2007. Challenge: peers on wheels—a road to new traffic information systems. In: *13th annual ACM Int.l conference on mobile computing and networking (MobiCom)*. New York, New York, USA.

Santa J, Moragon A, Gomez-Skarmeta AF. 2008. Experimental evaluation of a novel vehicular communication paradigm based on cellular networks. In: *IEEE intelligent vehicles symposium*. Eindhoven, Netherlands.

Seskar I, Marie S, Holtzman J, Wasserman J. 1992. Rate of location area updates in cellular systems. In: *IEEE vehicular technology conference (VTC '92)*. Denver, Colorado, USA.

Shinkawa T, Terauchi T, Kitani T, Shibata N, Yasumoto K, Ito M, Higashino T. 2006. A technique for information sharing using inter-vehicle communication with message ferrying. In: *IEEE Int.'l conference on mobile data management (MDM '06)*. Nara, Japan.

Sip2Peer. 2014. sip2peer. *Available at http://code.google.com/p/sip2peer*.

Stojmenovic I. 2008. Simulations in wireless sensor and ad hoc networks: matching and advancing models, metrics, and solutions. *IEEE Communications Magazine* **46(12)**:102–107 DOI 10.1109/MCOM.2008.4689215.

Sun Microsystems, Inc. 2005. JXTA technology: creating connected communities. In: *White paper*. Santa Clara: Sun Microsystems.

Tonguz O, Wisitpongphan N, Bai F, Mudalige P, Sadekar V. 2007. Broadcasting in VANET. In: *Workshop on vehicular ad hoc networks (Move'07)*. Montreal, Canada.

Trieber M, Hennecke A, Helbing D. 2000. Congested traffic states in empirical observations and microscopic simulations. *Physical Review E* **62(2)**:1805–1824 DOI 10.1103/PhysRevE.62.1805.

Vakulenko A. Circles intersection library for MATLAB. *Available at http://www.mathworks.com/matlabcentral/fileexchange/5313*.

Wegener A, Hellbruck H, Fischer S, Schmidt C, Fekete S. 2007. AutoCast: an adaptive data dissemination protocol fro traffic information systems. In: *66th IEEE vehicular technology conference (VTC '07)*. Dublin, Ireland.

Wi-Fi Alliance. 2015. Wi-Fi Direct standard. *Available at http://www.wi-fi.org/Wi-Fi_Direct.php*.

Xiang Q, Chen X, Kong L, Rao L, Liu X. 2015. Data preference matters: a new perspective of safety data dissemination in vehicular ad hoc networks. In: *IEEE INFOCOM*. IEEE.

Xu H, Barth M. 2006. An adaptive dissemination mechanism form inter-vehicle communication-based decentralized traffic information system. In: *9th Int'l IEEE conference on intelligent transportation systems (ITCS '06)*. Toronto, Canada.

Yan T, Zhang W, Wang G. 2014. DOVE: data dissemination to a desired number of receivers in VANET. *IEEE Transactions on Vehicular Technology* **63(4)**:1903–1916 DOI 10.1109/TVT.2013.2287692.

Zhao J, Cao G. 2007. VADD: Vehicle-assisted data delivery in vehicular ad hoc networks. *IEEE Transaction on Vehicular Technology* **57(3)**:1910–1922 DOI 10.1109/TVT.2007.901869.

Zhou B, Xu K, Gerla M. 2004. Group and swarm mobility models for ad hoc network scenarios using virtual tracks. In: *IEEE military communications conference (MILCOM)*. Monterey, California, USA.