POLITECNICO DI TORINO
Repository ISTITUZIONALE

Optimal Content Downloading in Vehicular Networks

(Article begins on next page)

30 April 2024

Optimal Content Downloading in Vehicular Networks

Francesco Malandrino, *Student Member, IEEE,* Claudio Casetti, *Member, IEEE,*
Carla-Fabiana Chiasserini, *Senior Member, IEEE,* and Marco Fiore, *Member, IEEE*

Abstract—We consider a system where users aboard communication-enabled vehicles are interested in downloading different contents from Internet-based servers. This scenario captures many of the infotainment services that vehicular communication is envisioned to enable, including news reporting, navigation maps and software updating, or multimedia file downloading. In this paper, we outline the performance limits of such a vehicular content downloading system by modelling the downloading process as an optimization problem, and maximizing the overall system throughput. Our approach allows us to investigate the impact of different factors, such as the roadside infrastructure deployment, the vehicle-to-vehicle relaying, and the penetration rate of the communication technology, even in presence of large instances of the problem. Results highlight the existence of two operational regimes at different penetration rates and the importance of an efficient, yet 2-hop constrained, vehicle-to-vehicle relaying.

Index Terms—Vehicular networks, downloading, optimization.

I. INTRODUCTION

The presence of high-end Internet-connected navigation and infotainment systems is becoming a reality that will easily lead to a dramatic growth in bandwidth demand by in-vehicle mobile users. Examples of applications of vehicular communication abound, and range from the updating of road maps to the retrieval of nearby points of interest, from the instant learning of traffic conditions to the download of touristic information and media-rich data files. This will induce vehicular users to resort to resource-intensive applications, to the same extent as today's cellular customers.

Most observers concur that neither the current nor the upcoming cellular technologies will suffice in the face of such a surge in the utilization of resource-demanding applications. Recent network overload episodes incurred in by cellular infrastructures in presence of smartphone users [1] provide a sobering wake-up call. To wit, a recent analysis on the traffic of a large US-based operator showed that smartphone users represent just 1% of the total subscribers, yet drain 60% of the network resources [2].

In order to design a network architecture that will scale to support the mass of vehicular users, one possibility is to offload part of the traffic to Dedicated Short-Range Communication (DSRC), through direct infrastructure-to-vehicle (I2V) transfer, as well as vehicle-to-vehicle (V2V) data relaying. Such an approach is especially attractive in the case of the download of large amounts of delay-tolerant data, a task that is likely to choke 3G/4G operator networks, but that well fits

DSRC-based I2V and V2V communication paradigms due to its lack of strict time constraints.

Within such a context, previous works on content downloading in vehicular networks have dealt with individual aspects of the process, such as the deployment of roadside Access Points (APs) [3]–[5], the performance evaluation of I2V communication [6], or the exploitation of specific V2V transfer paradigms [7], [8]. None of them, however, has tackled the problem as a whole, trying to quantify the actual potential of an I2V/V2V-based content downloading. In this paper, we identify the downloading performance limits achievable through DSRC-based I2V/V2V communication.

To this end, we assume ideal conditions from a system engineering viewpoint, i.e., the availability of preemptive knowledge of vehicular trajectories and perfect scheduling of data transmissions, and we cast the downloading process to a mixed integer linear programming (MILP) max-flow problem. The solution of such a problem yields the optimal AP deployment over a given road layout, and the optimal combination of any possible I2V and V2V data transfer paradigm. It thus represents the theoretical upper bound to the downloading throughput, under the aforementioned assumptions.

While it is true that the resulting problem is NP-complete, we show that, with a careful design of the model, it can be solved in presence of realistic vehicle mobility in a real-world road topology. In addition, we propose a sampling-based technique that efficiently yields a solution even for large-scale instances. Although the problem formulation and the performance figures we derive are interesting per-se, we also exploit our optimal solution to discuss the impact of key factors such as AP deployment, transfer paradigms and technology penetration rate.

As a final remark, we stress that our model, the first of its type to our knowledge, targets the general case of users interested in best-effort downloading of *different* data content. As a consequence, the goal is not to study information dissemination or cooperative caching, but to investigate the performance of content downloading.

This paper is organized as follows. Sec. II discusses previous work, while Sec. III describes the network scenario and our objectives. In Sec. IV, we build the graph modeling the vehicular network, while we formulate the max-flow problem in Sec. V. Results, derived in the scenarios described in Sec. VI, are presented in Sec. VII. In Sec. VIII, we evaluate the impact of our assumptions on the physical and MAC layers through ns-3 simulations. Finally, Sec. IX summarizes our major findings and points out directions for future research.

II. RELATED WORK

Our work relates to infrastructure deployment and content delivery in mobile environments, as well as to delay tolerant

networks. Below, we review the studies that are most relevant to ours, highlighting the novelty of our approach.

Infrastructure deployment. Earlier studies [9], [10] focus on the feasibility of using IEEE 802.11 APs to inject data into vehicular networks, as well as on the connectivity challenges posed by such an environment. In [11], the authors show that a random distribution of APs over the street layout can help routing data within urban vehicular ad hoc networks. In [12], the impact of several AP deployments on delay-tolerant routing among vehicles is studied. More precisely, each AP is employed as a static cache for content items that have to be transferred between vehicles visiting the AP at different times. Other than in the scope, the works in [11], [12] differ from ours also because they do not provide theoretical justification of the AP placements they propose.

AP deployment is formulated as an optimization problem in [13], where, however, the objective is not content downloading but the dissemination of information to vehicles in the shortest possible time. The study in [14], instead, estimates the minimum number of infrastructure nodes to be deployed along a straight road segment so as to provide delay guarantees to the data traffic that vehicles have to deliver to the infrastructure, possibly with the help of relays. A similar problem is addressed in [15], with the aim to support information dissemination. The different objectives of the above studies lead to completely different formulations, thus to results not comparable with the ones we present.

In [3], [4], infrastructure placement strategies are proposed that maximize the amount of time a vehicle is within radio range of an AP. Although longer periods of time under coverage can undoubtedly favor the download of contents by vehicular users, important differences with our work exist. First, our analysis is not limited to direct transfers from APs to vehicles, but includes traffic relaying. Second, while the problem formulation in [3] guarantees a minimum coverage requirement and the one in [4] maximizes the minimum-contact opportunity, we optimize the actual throughput, accounting for the airtime conflicts deriving from the contemporary presence of an arbitrary number of vehicles. Third, instead of studying a predefined set of paths over a given topology we process complete mobility traces.

An AP deployment strategy designed to favor content download through relaying in vehicular networks is introduced in [5]. The proposed optimization problem, however, aims at maximizing a metric reflecting the amount of vehicular traffic that enables V2V communication, and not the actual throughput. Moreover, such a formulation cannot capture the mutual interference among concurrent traffic transfers.

Content downloading and dissemination. With regard to content downloading in vehicular networks, the study in [16], unlike ours, focuses on the access to Web search and presents a system that makes such a service highly efficient by exploiting prefetching. Experimental and analytical results show the contribution of V2V and I2V communications to the system performance. The works in [7], [17] address the benefits of prefetching jointly with traffic scheduling techniques. In particular, the objective of [17] is to maximize the amount of data downloaded by vehicles through APs that form a wireless mesh network, given the AP deployment and an (imprecise) knowledge of the vehicles trajectory and of their connectivity with the APs. However, no multihop data transfer are investigated. In [7], both I2V and V2V communications are considered and the performance evaluation is carried out through simulation and a testbed on a circular campus bus route. Furthermore, a comparison against the solution to a max-flow problem is presented, but (i) it is limited to a simplified, highway-like scenario featuring one AP and one downloader and (ii) it assumes atomic contacts between nodes, hence neglecting interference and channel contention.

Our study also relates to cooperative downloading in vehicular networks. In this context, the work in [18], [19] introduces a vehicular peer-to-peer file sharing protocol, which allows vehicles to share a content of common interest. Our study on content download, instead, works in the more generic case where each user is interested in a different file. System assumptions similar to the ones made in [18], [19] are behind the works in [20], [21], about which, as a consequence, the same considerations hold.

Delay tolerant networks (DTNs). The vehicular cooperation paradigm that we consider relates our work to DTNs. In particular, [22] assesses the benefit to content dissemination of adding varying numbers of base stations, mesh nodes and relay nodes to a DTN, through both a real testbed and an asymptotic analysis. A DTN time-invariant graph, which is similar to the time-expanded graph used in our study, was presented in [23]. With respect to [23], we do not assume the contacts between mobile nodes to be atomic but to have arbitrary duration, and we build the network graph so as to account for the presence of roadside infrastructure and channel contention. The representation of a time-varying network topology as a time-expanded graph can be found in [24], [25], where the former is an earlier version of this work. As for the latter, such a graph is used to identify the nodes whose limited storage may impair the network performance, and to formulate a max-flow problem whose solution leads to an optimal, distributed routing and storage policy. In our work, we address the performance limits of content downloading and the problem of AP deployment, for which no distributed solution is needed.

III. NETWORK SYSTEM AND GOALS

We envision a network composed of fixed roadside APs and vehicular users, where some of the latter (hereinafter named *downloaders*) are interested in downloading best-effort traffic from the Internet through the APs. We consider the general case in which every downloader is interested in different content: downloaders can either exploit direct connectivity with the APs, if available, or be assisted by other vehicles acting as intermediate relays. Specifically, we account for all possible data *transfer paradigms* that can be implemented through I2V/V2V communication:

direct transfer, resulting from a direct communication between an AP and a downloader. This represents the typical way mobile users interact with the infrastructure in today's wireless networks;

connected forwarding, i.e., traffic relaying through one or more vehicles that create a multi-hop path between an AP

and a downloader, where all the links of the connected path exist at the time of the transfer. This is the traditional approach to traffic delivery in ad hoc networks;

carry-and-forward, i.e., traffic relaying through one or more vehicles that store and carry the data, eventually delivering them either to the target downloader or to another relay deemed to meet the downloader sooner.

We stress that connected forwarding and carry-and-forward are inherently multi-hop paradigms. We assume that vehicular users are rational, hence they can be engaged in relaying traffic for others only if they are not currently retrieving the content for themselves. Furthermore, since our goal is to derive an upper bound to the system performance, we assume the availability of preemptive knowledge of vehicular trajectories and perfect scheduling of data transmissions.

From the viewpoint of the network system, we consider that each node (a vehicle or an AP) has one radio interface only. This is a common assumption for vehicular nodes, while the extension to the case where APs have more than one interface is straightforward. Any two nodes in the network can communicate at a given time instant, i.e., they are neighbors, if their distance is below or equal to their maximum radio range. Also, we assume that the maximum radio range is common to all network nodes and is equal to the node interference range[1]. We consider that V2V communications occur on the same frequency channel, which is different from the channels used for I2V communication[2]; APs with overlapping coverage areas operate on separate channels. When under AP coverage, a vehicle can always choose either I2V or V2V communication. The nodes share the channel bandwidth allocated for service applications using an IEEE 802.11-based MAC protocol.

Our objective is to design the content downloading system so as to maximize the aggregate throughput. To this aim, we have to jointly solve two problems: (i) given a set of candidate locations and a number of APs to be activated, we need to identify the deployment yielding the maximum throughput; (ii) given the availability of different data transfer paradigms, possibly involving relays, we have to determine how to use them in order to maximize the data flow from the infrastructure to the downloaders. Our approach consists in processing a road layout and an associated vehicular mobility trace, so as to build a graph that represents the temporal network evolution (Sec. IV). By using this graph, we formulate a max-flow problem whose solution matches our goals (Sec. V).

IV. DYNAMIC NETWORK TOPOLOGY GRAPH

We generate a time-expanded graph [26], hereinafter dynamic network topology graph (DNTG), from a vehicular mobility trace. To build the graph, we consider that on the road layout corresponding to the mobility trace there are: (i) a set of A candidate locations (a_i, $i = 1, \ldots, A$) where APs could be placed, (ii) a set of V vehicles (v_i, $i = 1, \ldots, V$) transiting over the road layout and participating in the network, and (iii)

a subset of D vehicles that wish to download data from the infrastructure.

The aim of the DNTG is to model all possible opportunities through which data can flow from the APs to the downloaders, possibly via relays. Given the mobility trace, we therefore identify the *contact events* between any pair of nodes (i.e., two vehicles, or an AP and a vehicle). Each contact event is characterized by:
(i) the quality level of the link between the two nodes. Several metrics could be considered; here, we specifically take as link quality metric the data rate achievable at the network layer;
(ii) the contact starting time, i.e., the time instant at which the link between the two nodes is established or the quality level of an already established link takes on a new value;
(iii) the contact ending time, i.e., the time instant at which the link is removed, or its quality level has changed.
We stress that, by associating a time duration to the contact events, instead of considering them as atomic, we can model critical aspects such as channel contention.

The time interval between any two successive contact events in the network is called *frame*. Within a frame the network is static, i.e., no link is created or removed and the link quality levels do not change. We denote by F the number of frames in the considered trace, and by τ^k the duration of the generic frame k ($1 \leq k \leq F$); also, all on-going contact events during frame k are said to be *active* in that frame.

Each vehicle v_i participating in the network at frame k is represented by a vertex v_i^k ($1 \leq i \leq V$) in the DNTG, whereas every candidate AP location a_i is mapped within each frame k onto a vertex a_i^k ($1 \leq i \leq A$). We denote by $\mathcal{V}^k$ and $\mathcal{A}^k$ the set of vertices representing, respectively, the vehicles and candidate AP locations in the DNTG at time frame k, while we denote by $\mathcal{D}^k \subseteq \mathcal{V}^k$ the subset of vertices representing the downloaders that exist in the network at frame k. All non-downloader vehicles in $\mathcal{R}^k = \mathcal{V}^k \setminus \mathcal{D}^k$ can act as relays, according to the data transfer paradigms outlined above.

Within each frame k, a directed edge (v_i^k, v_j^k) exists from vertex $v_i^k \in \mathcal{R}^k$ to vertex $v_j^k \in \mathcal{V}^k$, if a contact between the non-downloader vehicle v_i and another vehicle v_j is active during that frame. Each edge of this type is associated with a weight $w(v_i^k, v_j^k)$, equal to the rate of that contact event. The set including such edges is defined as $\mathcal{L}_v^k$. Similarly, a directed edge (a_i^k, v_j^k) exists from vertex $a_i^k \in \mathcal{A}^k$ to vertex $v_j^k \in \mathcal{V}^k$, if a contact between the candidate AP location a_i and the vehicle v_j is active during frame k. Again, these edges are associated with weights $w(a_i^k, v_j^k)$, corresponding to the contact event rate, and their set is defined as $\mathcal{L}_a^k$.

A directed edge (v_i^k, v_i^{k+1}) is also drawn from any vertex $v_i^k \in \mathcal{R}^k$ to any vertex $v_i^{k+1} \in \mathcal{R}^{k+1}$, for $1 \leq k < F$. While the edges in $\mathcal{L}_v^k$ and $\mathcal{L}_a^k$ represent transmission opportunities, those of the form (v_i^k, v_i^{k+1}) model the possibility that a non-downloader vehicle v_i physically carries some data during its movement in the time interval from frame k to frame $k + 1$. Accordingly, these edges are associated with a weight representing the vehicle memory capabilities, since they do not imply any rate-limited data transfer over the wireless medium. However, dealing with vehicular nodes as opposed to resource-constrained hand-held devices, we assume the weight

---

[1]Although simplistic, the impact of such an assumption on the system performance is negligible, as shown by the comparison between analytical and simulation results in Sec. VIII.

[2]Single-radio multichannel management is foreseen by current standardization activities on vehicular communication systems, e.g., IEEE 1609.4.
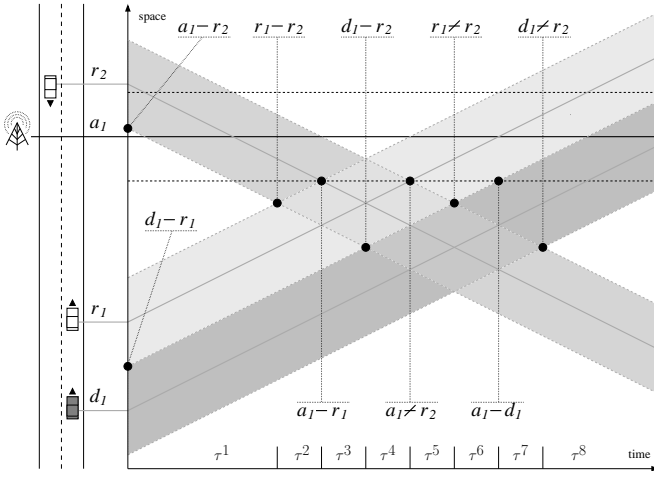
(a) Space-time representation of the contact events
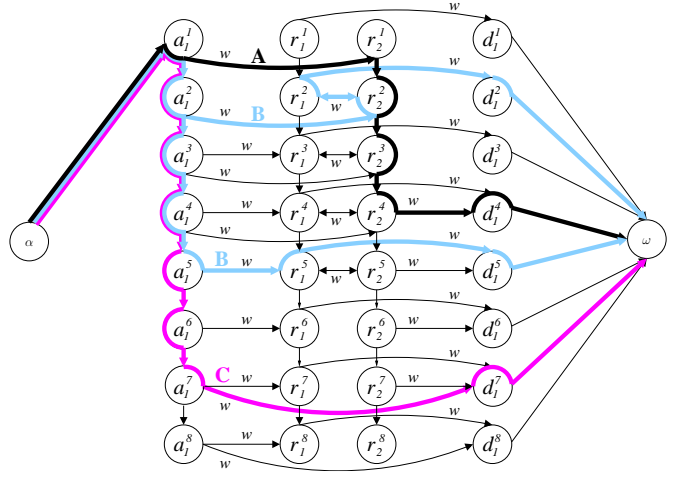
(b) DNTG resulting from the contact events

Fig. 1. A sample set of contact events (a) and the corresponding DNTG (b), in presence of one candidate AP location and three vehicles, the first of which (d_1) is a downloader while the others (r_1, r_2) can act as relays. In (a), shadowed areas represent *halved* transmission ranges, so that links exist when two shadowed areas touch or overlap, and break when such areas become disjoint. These events allow to fragment the time into frames of duration $\tau^1, \ldots, \tau^8$ (for simplicity, here the link quality is assumed constant). The network connectivity during each frame is represented by a row of vertices in the DNTG. In the graph, we highlight paths that are representative of the carry-and-forward (A), connected forwarding (B), and direct (C) transfer paradigms

of such edges to take on an infinite value. A directed edge (a_i^k, a_i^{k+1}) of infinite weight is also drawn between any two vertices representing the same candidate AP location at two consecutive frames, i.e., from $a_i^k \in \mathcal{A}^k$ to $a_i^{k+1} \in \mathcal{A}^{k+1}$ ($1 \leq k < F$). We will refer to the edges of the kind (v_i^k, v_i^{k+1}) or (a_i^k, a_i^{k+1}) as intra-nodal.

Finally, in order to formulate a max-flow problem over the DNTG, we introduce two virtual vertices, α and ω, respectively representing the source and destination of the total flow over the graph. Then, the graph is completed with infinite-weight edges (α, a_i^1), from α to any vertex $a_i^1 \in \mathcal{A}^1$, and (v_i^k, ω), from any vertex $v_i^k \in \mathcal{D}^k$ to ω, $1 \leq k \leq F$.

The DNTG is therefore a weighted directed graph, representing the temporal evolution of the network topology. An example of its derivation is given in Fig. 1, in presence of one AP location and three vehicles d_1, r_1, r_2, with d_1 being a downloader and r_1, r_2 possibly acting as relays. Fig. 1(a) depicts the spatio-temporal evolution of node positions: there, contact events are highlighted through the times at which links are established or lost. For simplicity, in this example we assume the achievable network-layer rate w to be constant during the entire lifetime of a link. The durations of the frames, within which the network connectivity is unchanged, are denoted by $\tau^1, \ldots, \tau^8$. In Fig. 1(b), frames correspond to rows of vertices in the DNTG, where intra-nodal edges connect vertices representing the same vehicle or candidate AP location over time. Note that the graph allows us to capture all the data transfer paradigms previously discussed. It is thus possible to identify paths in the graph that correspond to (i) direct download from the candidate AP to the downloader, as path C, (ii) connected forwarding through 3-hops (frame 2) and 2-hops (frame 5), as path B, and (iii) carry-and-forward through the movement in time of the relay r_2, as path A.

V. THE MAX-FLOW PROBLEM

Given the DNTG, our next step is the formulation of an optimization problem whose goal is to maximize the flow from

α to ω, i.e., the total amount of downloaded data. Denoting by $x(\cdot, \cdot)$ the traffic flow over an edge connecting two generic vertices, our objective can be expressed as:

$$\max \sum_{k=1}^{F} \sum_{v_i^k \in \mathcal{D}^k} x(v_i^k, \omega). \quad (1)$$

The max-flow problem needs to be solved taking into account several constraints, listed below.

A. Constraints

Non-negative flow. The flow on every existing edge must be greater than or equal to zero.

Flow conservation. For any vertex in the DNTG, the amount of incoming flow must equal the amount of outgoing flow. This constraint is expressed in slightly different form, depending on whether the vertex represents a downloader, a relay, or a candidate AP location. For the generic vertex representing a downloader, $v_i^k \in \mathcal{D}^k$, and any frame k, this maps onto:

$$\sum_{\substack{a_j^k \in \mathcal{A}^k: \\ (a_j^k, v_i^k) \in \mathcal{L}_a^k}} x(a_j^k, v_i^k) + \sum_{\substack{v_j^k \in \mathcal{R}^k: \\ (v_j^k, v_i^k) \in \mathcal{L}_v^k}} x(v_j^k, v_i^k) = x(v_i^k, \omega). \quad (2)$$

For any frame k and potential relay vertex, $v_i^k \in \mathcal{R}^k$:

$$\sum_{\substack{a_j^k \in \mathcal{A}^k: \\ (a_j^k, v_i^k) \in \mathcal{L}_a^k}} x(a_j^k, v_i^k) + \sum_{\substack{v_j^k \in \mathcal{R}^k: \\ (v_j^k, v_i^k) \in \mathcal{L}_v^k}} x(v_j^k, v_i^k) + \mathbb{1}_{\left[(v_i^{k-1}, v_i^k)\right]} x(v_i^{k-1}, v_i^k)$$

$$= \sum_{\substack{v_j^k \in \mathcal{V}^k: \\ (v_i^k, v_j^k) \in \mathcal{L}_v^k}} x(v_i^k, v_j^k) + \mathbb{1}_{\left[(v_i^k, v_i^{k+1})\right]} x(v_i^k, v_i^{k+1}) \quad (3)$$

where the indicator function is equal to 1 if the specified edge exists, and it is 0 otherwise.

For each vertex representing a candidate AP, $a_i^k \in \mathcal{A}^k$:

$$\mathbb{1}_{[k=1]}x(\alpha, a_i^k) + \mathbb{1}_{[k>1]}x(a_i^{k-1}, a_i^k) =$$
$$\mathbb{1}_{[k<F]}x(a_i^k, a_i^{k+1}) + \sum_{\substack{v_j^k \in \mathcal{V}^k: \\ (a_i^k, v_j^k) \in \mathcal{L}_a^k}} x(a_i^k, v_j^k) \quad (4)$$

where the indicator functions are equal to 1 if the specified condition holds, and 0 otherwise. Note that, for a vertex representing a candidate AP location, ingoing flows may come from a vertical edge or from α, while outgoing flows may be over a vertical edge, or over edges toward vehicle vertices.

Finally, we impose that the total flow exiting α equals the total flow entering ω: $\sum_{a_i^1 \in \mathcal{A}^1} x(\alpha, a_i^1) = \sum_{k=1}^F \sum_{v_j^k \in \mathcal{D}^k} x(v_j^k, \omega)$.

Channel access. As mentioned, we deal with unicast transmissions and assume that the nodes use a 802.11-based MAC scheme; also, V2V and I2V communications occur on different channels. Then, given a tagged vehicle, we consider that none of the following events can take place simultaneously, and the time span of each frame must be shared among them:

1) the vehicle transmits to a neighboring vehicle;
2) a neighboring vehicle receives from any relay;
3) the vehicle receives from a neighboring relay;
4) a neighboring relay transmits to any vehicle;
5) the vehicle receives from a neighboring AP.

Recall that we do not model the scheduling of the single packets transmitted within each frame. Rather, we consider the total amount of data carried by each flow. Also, in 2) a neighboring vehicle receiving data is accounted for; in presence of hidden terminals, this still holds if the RTS/CTS handshake is used. Considering that: 1) is a subcase of 2) and 3) is a subcase of 4), for the generic vertex $v_i^k \in \mathcal{V}^k$ and for any frame k, we have:

$$\sum_{\substack{v_j^k \in \mathcal{R}^k, v_m^k \in \mathcal{V}^k \\ (v_j^k, v_m^k) \in \mathcal{L}_v^k}} \mathbb{1}_{\left[(v_m^k, v_i^k)||(v_j^k, v_i^k)\right]} \frac{x(v_j^k, v_m^k)}{w(v_j^k, v_m^k)} + \sum_{\substack{a_j^k \in \mathcal{A}^k: \\ (a_j^k, v_i^k) \in \mathcal{L}_a^k}} \frac{x(a_j^k, v_i^k)}{w(a_j^k, v_i^k)} \leq \tau^k \quad (5)$$

where the indicator functions are defined as before.

In addition, for each candidate AP, we have that its total transmission time during the generic frame k cannot exceed the frame duration. Thus, for any k and $a_j^k \in \mathcal{A}^k$, we have:

$$\sum_{\substack{v_i^k \in \mathcal{V}^k: \\ (a_j^k, v_i^k) \in \mathcal{L}_a^k}} \frac{x(a_j^k, v_i^k)}{w(a_j^k, v_i^k)} \leq \tau^k . \quad (6)$$

The previous constraints allow a vehicle under coverage of an AP to use I2V and V2V communications within the same frame. Next, we consider the case where a vehicle under the coverage of (at least) one AP is not configured to operate in ad hoc mode, i.e., it cannot communicate with other vehicles. Then, for any frame k and $v_j^k \in \mathcal{R}^k, v_m^k \in \mathcal{V}^k$ such that $(v_j^k, v_m^k) \in \mathcal{L}_v^k$, the following constraint holds:

$$x(v_j^k, v_m^k) \leq \left(1 - \max_{\substack{a_i^k \in \mathcal{A}^k: \\ (a_i^k, v_j^k)||(a_i^k, v_m^k) \in \mathcal{L}_a^k}} \{y_i\}\right) w(v_j^k, v_m^k)\tau^k \quad (7)$$

where y_i, $i = 1, \ldots, A$, are Boolean variables, whose value is 1 if an AP is placed at candidate location i and 0 otherwise. If there is at least an AP within the vehicle range, the first term of the product becomes 0, thus imposing that the flow on all edges $(v_j^k, v_m^k) \in \mathcal{L}^k$ is 0.

Overlapping AP coverages. Recall that, when a vehicle falls within coverage of two or more APs, we assume that, during a frame, it communicates with one AP only, and that the APs operate on different frequency channels. We therefore introduce a second set of Boolean variables t_{ij}^k ($1 \leq i \leq A$, $1 \leq j \leq V$, $1 \leq k \leq F$) whose value is 1 if the candidate AP a_i communicates with the vehicle v_j during frame k and 0 otherwise. Then, for every candidate AP vertex $a_i^k \in \mathcal{A}^k$, vehicle vertex $v_j^k \in \mathcal{V}^k$, and frame k, we impose that

$$t_{ij}^k \in \{0, 1\} \ ; \ \sum_{i=1}^A t_{ij}^k \leq 1 \ ; \ x(a_i^k, v_j^k) \leq w(a_i^k, v_j^k)\tau^k t_{ij}^k .$$

Maximum number of APs. The last set of constraints imposes that no more than $\hat{A}$ candidate AP locations are selected, through the variables y_i. Then, for any i, we write:

$$y_i \in \{0, 1\} \ ; \ \sum_{i=1}^A y_i \leq \hat{A} \ ; \ x(\alpha, a_i^1) \leq My_i$$

where $M \in \mathbb{R}$ is an arbitrarily large positive constant.

B. Modeling the transfer paradigms

We now describe how the different transfer paradigms introduced in Sec. III are modeled in our formulation.

The traffic transferred through the direct paradigm corresponds to the amount of data that, at any frame k, flows from one candidate AP vertex, $a_i^k \in \mathcal{A}^k$, to $v_j^k \in \mathcal{D}^k$. As for the traffic transferred through connected forwarding, this is represented as the amount of data that, at any frame k, flows from one relay vertex $v_i^k \in \mathcal{R}^k$ to a downloader vertex, with one or more edges connecting v_i^k to the vertex representing the AP that originated the data. Such a situation indeed corresponds to the case where a multi-hop connected path between an AP and a downloader exists. The data transferred through carry-and-forward, instead, correspond to the flow associated with any (v_i^k, v_j^k) edge at frame k (with $v_i^k \in \mathcal{R}^k, v_j^k \in \mathcal{D}^k$), such that the relay vertex v_i^k is no longer connected (either directly or through multiple edges) to the vertex representing the AP that originated the flow.

Furthermore, while deriving the results, we consider three possible cases. In the *unlimited* case, no limitation is imposed to the maximum number of relays used to deliver traffic to a downloader. This is modeled simply using the constraints listed in Sec. V-A. In the *2-hop limit* case, at most one relay can be employed. This is studied by imposing that transmissions between relays cannot occur, i.e., $x(v_i^k, v_j^k) = 0$ for $1 \leq k \leq F$ and $v_i^k, v_j^k \in \mathcal{R}^k$, such that $(v_i^k, v_j^k) \in \mathcal{L}_v^k$. In the *1-hop limit* case, only 1-hop transfers from an AP to a downloader are allowed; we represent this case by imposing that $x(v_i^k, v_j^k) = 0$ for any k and $v_i^k \in \mathcal{R}^k, v_j^k \in \mathcal{V}^k$ such that $(v_i^k, v_j^k) \in \mathcal{L}_v^k$.

Further refinements on the representation of the transfer paradigms can be found in Appendix A-1 (see Supplemental Material).

C. Sampling-based solution

The problem falls in the category of mixed integer linear programming (MILP) problems. Denoting by T the average vehicle trip duration in frames and being VT/F the average number of vehicles in the road layout at a given instant, the number of decision variables is $O\left((VT)^2/F\right)$. The number of constraints is of the same order of magnitude as the one of the number of variables. A more detailed discussion on the problem complexity can be found in Appendix A-2.

We solve the problem through the Gurobi solver, which uses a variant of the branch-and-cut algorithm. However, due to the large number of constraints involving Boolean variables, solving the MILP on the full DNTG is impractical for large instances (e.g., large geographical areas, high number of vehicles participating in the content downloading, or large number of candidate AP locations). To be able to analyze such cases, we resort to a graph sampling approach. More specifically, we take the following steps:
1) we sample the DNTG obtaining a small, yet representative, sub-graph, which includes all relevant candidate AP locations (as detailed below);
2) we find the optimal AP deployment using such a sub-graph;
3) we apply the obtained deployment to the full graph and optimize the flows – a linear programming (LP) problem that can be easily solved as it does not involve Boolean variables.

In order to accomplish the first step, the selected sub-graph must include the α and ω vertices, and reflect the characteristics of the original graph (e.g., the relevance of the candidate AP locations). Since we have to collect not only a representative set of the graph vertices, but a *connected* sample, uniform vertex sampling is not a viable option. Thus, we resort to a random walk-based approach [27], and devise a tailored variant of it. Such a variant is needed to effectively cope with the following challenging peculiarities that our DNTG exhibits with respect to ordinary graphs.

First, not only is the DNTG directed, but the flow goes from α to ω, while the edges are specifically directed from candidate AP location vertices to vehicle vertices, as well as from lower to higher values of the frame index k. This implies that it is not possible to make an arbitrarily long walk on the DNTG; thus, we need to combine vertices and edges that are sampled over subsequent multiple short walks.

Second, while walking from α to ω would be a natural choice in ordinary graphs, in our case this would turn into sampling relay vehicle vertices that may not be connected with downloaders, hence with ω (see top Fig. 2). To avoid the unnecessary sampling of these vertices, we let the walks go from ω to α, crossing each edge along its opposite direction. Each walk therefore goes from ω to one or more vertices representing one downloader, then possibly to relay vertices, to one candidate AP location and finally to α (see bottom Fig. 2). An example is shown in Fig. 2(d). Note that, by adopting such a strategy, we obtain a fairly small subgraph,
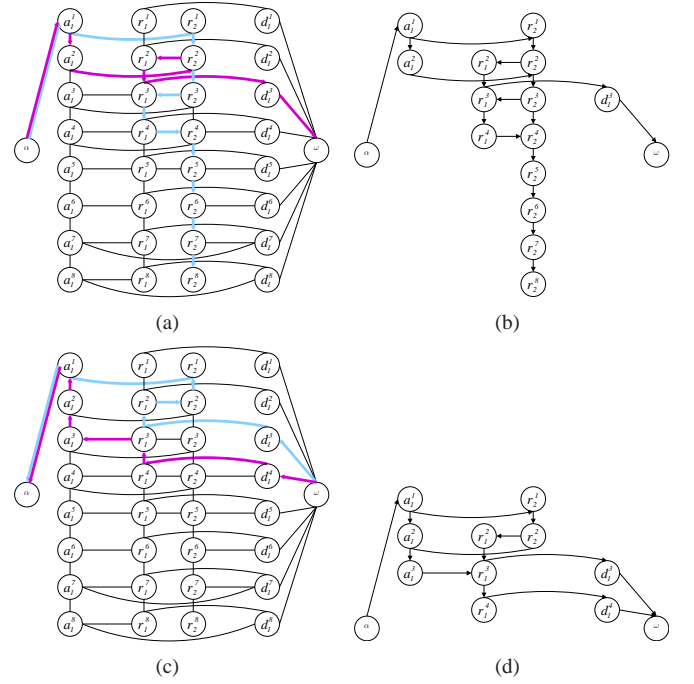

Fig. 2. Example of a DNTG sampling when walks cross each edge along its direction (a) or its opposite direction (c). Resulting sampled graph are in (b) e (d), respectively. Arrows refer to the walk direction (left figures) and to the edge direction (right figures)

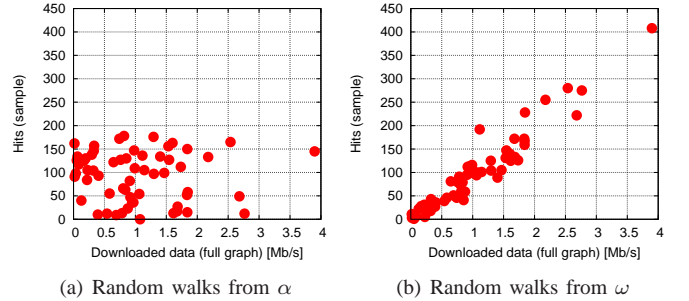(a) Random walks from α (b) Random walks from ω

Fig. 3. Number of times that a vertex corresponding to a candidate AP location is sampled, vs. the amount of data per second downloaded through that AP in the full DNTG. Walks start from α in (a) and from ω in (b)

yet containing vertices representing several vehicles (including relays and downloaders), as well as candidate AP locations.

Another desirable effect of the above strategy is that the candidate AP locations and the relays are sampled with a frequency that is proportional to the number of paths between α and ω passing through them, while the downloaders are sampled with a frequency that is proportional to their trip duration. We support such a statement by looking at the correlation between the relevance of candidate AP locations for content downloading, and the number of walks including each candidate AP location. The relevance is expressed as the amount of data per second outgoing from each candidate AP location under the max-flow solution in the full DNTG, in the scenario in Fig. 4(a) with $\hat{A} = 60$. As is clear from Fig. 3(b), there is a strong correlation when walks start from ω; conversely, with the standard sampling (i.e., for walks starting from α) there is no evident correlation (Fig. 3(a)).

The performance of the max-flow solution over the sampled DNTG is compared to that over the full graph in Sec. VII-A.

(a) Urban: traffic over space (b) Village: traffic over space (c) Suburban: traffic over space (d) Network-layer rate
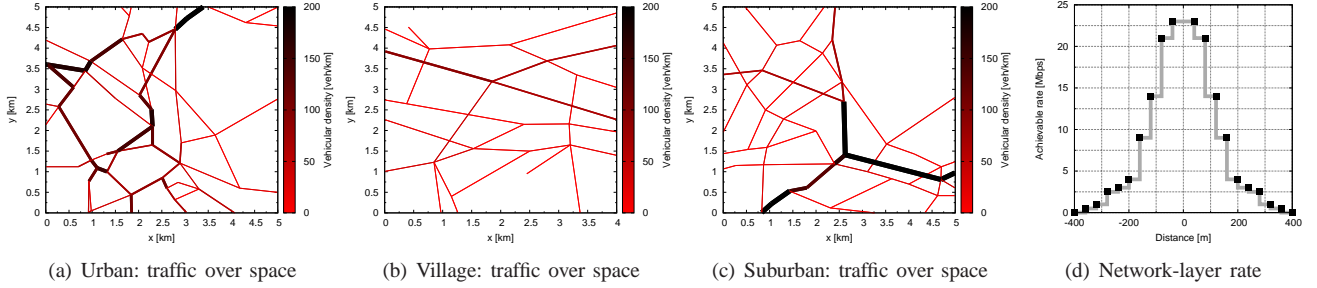
Fig. 4. Road layout in the urban (a), village (b) and suburban (c) scenarios, and achievable network-layer rate characterization as a function of distance (d)

VI. REFERENCE SCENARIOS

We consider real-world road topologies representing different environments, namely the urban area of Zurich, the village area of Schlieren and the suburban area of Wallisellen, in Switzerland. Each road topology covers an area of 20 km^2; the vehicular mobility in the region has been synthetically generated at ETH Zurich [28]. The macroscopic- and microscopic-level models employed to produce the movement traces allow a realistic representation of the vehicular mobility, in terms of both large-scale traffic flows and small-scale V2V interactions. Although our model can accommodate any frame duration, so as to reflect, e.g., faster variations of the link quality, given the 1-second time granularity of the trace, we consider $\tau^k \geq 1$ s ($k = 1, \ldots, F$).

Since we use a realistic mobility model, in each road topology the vehicular traffic intensity varies depending on the road segment and time of the day. In Figs. 4(a)–4(c), we report the road layout of the urban, village and suburban environments, highlighting the different traffic volumes observed over each road segment: thicker, darker segments identify the roads characterized by higher vehicular density. As far as vehicular traffic variations over time are concerned, we consider only time periods corresponding to medium-high vehicle density. In the urban, village and suburban traces, each lasting about 5 hours, this leads to an average density of 90, 62.5 and 33.5 veh/km, respectively. Further details on the behavior of vehicular traffic and on the selection of candidate AP locations can be found in Appendix A-3.

We consider different values of the technology penetration rate, i.e., the fraction of vehicles equipped with a communication interface and willing to participate in the content downloading process; we denote such a parameter by p. Also, the percentage of such communication-enabled vehicles that concurrently request content, i.e., that act as downloaders, is denoted by d. Unless otherwise specified, we will consider $d = 0.01$ (i.e., 1% of the vehicles participating in the network) – a reasonable value as observed in wired networks [29].

The value of the achievable network-layer rate between any two nodes is adjusted according to the distance between them. To this end, we refer to the 802.11a experimental results in [6] to derive the values shown in Fig. 4(d), and we use them as samples of the achievable network-layer rate. Note that we limit the maximum node transmission range to 200 m, since, as stated in [6], this distance allows the establishment of a reliable communication in 80% of the cases.

Given the above settings and that $\hat{A}$ APs have to be deployed, in the next section we present the performance obtained by solving the max-flow problem on the full DNTG, or on its sampled version. We thus attain the optimal AP deployment as well as the values of the flow variables corresponding to the amount of data that downloaders receive. Using the flow values, we can then compute: (i) the per-user throughput, as the ratio of the amount of received data to the downloader trip duration; (ii) the fraction of traffic delivered through the direct, connected forwarding, or carry-and-forward paradigm; (iii) the Jain's fairness index, computed on the average throughput obtained by each downloader; (iv) the average packet delivery delay from AP to downloader, accounting for both I2V and V2V communication.

Our problem formulation can also accommodate any specific AP deployment by fixing the values of the binary variables y_i. The system throughput is then obtained as the output of the max-flow problem given the selected AP locations y_i. Solving the max-flow problem implies the optimization of the traffic scheduling, i.e., the values taken by the V2V and I2V flow variables at any frame k ($x(v_j^k, v_i^k)$, $x(a_j^k, v_i^k)$). The results thus represent the best performance achievable under the chosen deployment and the assumptions made in Sec. III.

We leverage the capability of our framework to model different AP deployments and explore the following strategies:
Random: $\hat{A}$ locations are randomly selected among the candidate ones, according to a uniform distribution;
Crowded: $\hat{A}$ locations are picked, whose coverage area exhibits, over time, the highest vehicular density;
Contact: $\hat{A}$ locations are selected, which maximize the sum of the contact opportunities between vehicles and APs. Inspired by the metric adopted in [4], for each vehicle we express the contact opportunity as the fraction of the road segment lengths traveled while under coverage of at least one AP.

VII. PERFORMANCE EVALUATION

In this section, we evaluate the performance of content downloading in vehicular networks, by assessing the impact of different settings on the system.

Specifically, in Sec. VII-A, we evaluate the impact that the penetration rate of the vehicular communication technology, p, has on the content downloading performance. Our results reveal the existence of two regimes, separating initial deployment stages (characterized by $p < 20\%$) from a mature technology (i.e., $p > 30\%$).

These two working regimes are analysed in detail in Sec. VII-B and Sec. VII-C, respectively. For each regime, we

discuss the impact of the AP deployment strategies, transfer paradigms and the road environment on the downloading performance. For the high-penetration regime, we also investigate the system behavior as the percentage of concurrent downloaders varies and in presence of overlapping AP coverages.

A. Impact of vehicular communication technology adoption

As a first step in the evaluation of vehicular content downloading, we look at the impact that the diffusion of I2V and V2V communication technologies has on the system performance. To that end, we consider different values of p as well as different extensions of the roadside AP deployment. For clarity, we focus on the urban scenario depicted in Fig. 4(a) and we consider the AP deployment obtained by solving the max-flow problem on the full and the sampled DNTG. Also, we study non-overlapping AP coverages and constrain V2V relaying to 2 hops from APs; these assumptions will be relaxed later on.

Fig. 5 portrays the evolution of the key performance metrics when the technology penetration rate, p, varies between 5% and 80%. The curves refer instead to different extensions of the roadside infrastructure, ranging from $\hat{A} = 15$ to $\hat{A} = 60$ APs. Note that the latter value essentially corresponds to a complete coverage of the road topology by the APs. The results obtained using full and sampled DNTG are denoted by thick and thin lines, respectively.

Throughput. The average per-downloader throughput, in Fig. 5(a), is very satisfying in all conditions, scoring well above 10 Mb/s even in low-p, low-$\hat{A}$ scenarios, and more than 20 Mb/s in presence of a wide I2V and V2V technology adoption. When separating the effects of $\hat{A}$ and p, the availability of a more pervasive (although non overlapping) infrastructure coverage helps at both low and high-penetration rates, although its impact is lower than one could expect. Indeed, a pervasive 60-AP deployment only results in a constant 3-Mb/s gain over a simple 15-AP deployment. Higher improvements can be instead obtained from the spread of in-car communication interfaces, with an average throughput increase of 8 Mb/s as p grows from 5% to 80%. We remark that the steepest throughput growth lies between 5% and 20% penetration rate.

Delay. Delays, in Fig. 5(b), are in the order of tens of seconds, a good result when considering the delay-tolerant nature of most V2V transfers. We can observe different behaviors for low and high values of p. For $p < 20\%$, an increase in availability of relays leads to more frequent V2V transfers, hence higher delays. When $p > 30\%$, the already pervasive presence of relays makes the impact of even higher penetration rates negligible. Also, a denser AP deployment helps reducing the delay, although such a gain is significant only when $\hat{A}$ is low.

Fairness. To get an insight on how the system throughput is actually shared among the downloaders, Fig. 5(c) shows the Jain's fairness index. The increase in penetration rate has a major impact since it implies a growing number of V2V communication opportunities. Indeed, for low values of p, downloaders travelling over secondary roads have fewer chances to benefit from traffic relay than downloaders travelling on main (typically, more crowded) roads. It follows that

some unfairness arises for low penetration rates, while the system becomes fair for medium-high values of p. Also, the larger the $\hat{A}$, the higher the level of fairness, as both main and secondary roads can be covered.

Transfer paradigm. The above observations on the fundamental role of V2V traffic relaying is confirmed by the results in Fig. 5(d), depicting the fraction of content downloaded through relay vehicles. Indeed, most of the content is received by downloaders from relays (through either connected forwarding or carry-and-forward). Clearly, the importance of V2V communication tends to grow with the penetration rate p, since the availability of additional relays allows a more intensive utilization of the wireless resources. More surprisingly, the presence of additional APs only marginally reduces the utilization of V2V communication, and, at high values of p, more than 80% of the data is downloaded through relays even when the whole road surface is covered by APs. We will further comment on this phenomenon later in this section.

Problem solution. Fig. 5 highlights the effectiveness of the sampling-based technique introduced in Sec. V-C, when compared against the optimization solution on the full DNTG. The performance results obtained with the latter are shown in all the plots as thick grey curves, while the thinner lines represent the outcome of the sampling-based solution. The throughput and delay loss induced by the sampling are negligible, and the fraction of V2V downloading is identical in the two cases. The only noticeable difference can be observed in terms of fairness, since, by sampling the vertices representing the candidate AP locations with higher weight, APs on secondary roads are seldom activated in the max-flow solution, thus reducing the level of fairness.

We remark that sampling the DNTG allows to solve significantly more complex instances of the max-flow problem. As an example, in the plots of Fig. 5, memory requirements become too demanding for the solution of the complete problem when more than 20% of the vehicles are part of the network.

Summary. The performance metrics are consistent in revealing the critical importance of the penetration rate p and the lower impact of the roadside infrastructure extension. Accordingly, we can separate two regimes. The first, when $p < 20\%$, i.e., at early stages of the technology adoption, characterized by lower throughput and higher delay, a stronger dependency on direct I2V communication and lower downloading fairness. The second, for $p > 30\%$, i.e., in presence of a quite mature technology, featuring instead higher throughput and lower delay, massive use of V2V communication and high fairness. As the impact of the system settings is different within these two regimes, in the following we will study them separately. According to the results above, we will employ the max-flow problem solution on the complete and on the sampled graph in the low- and high-penetration regime, respectively.

B. Low-penetration regime

As case study of the low-penetration regime, we consider $p = 10\%$. The default settings include the urban scenario, non-overlapping AP coverages, a 1% fraction of downloaders and a 2-hop limit in V2V relaying.
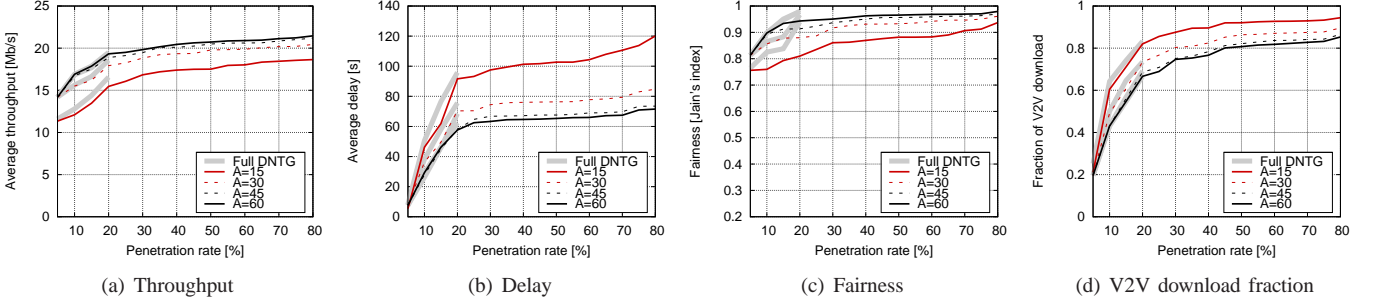
Fig. 5. Max-flow strategy: Average per-downloader throughput (a), delay (b), fairness (c), and V2V downloading fraction (d) versus the penetration rate
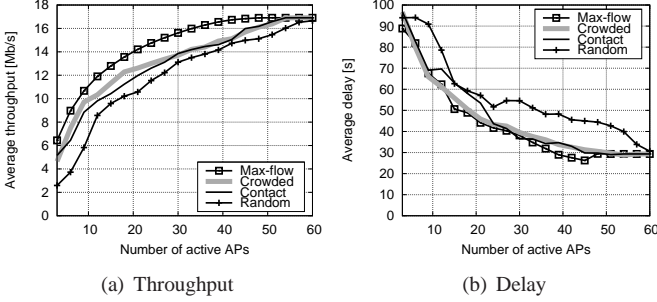

Fig. 6. Low-penetration regime: Average per-downloader throughput (a) and delay (b), vs. $\hat{A}$, for different AP deployment strategies
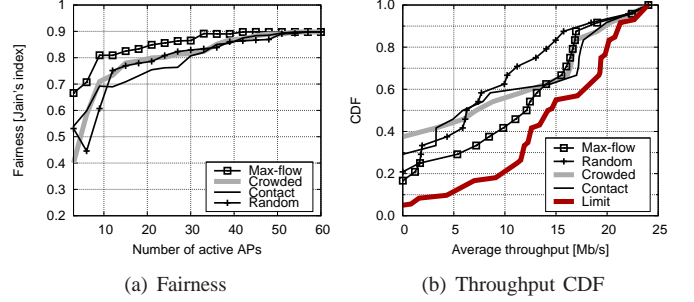
Fig. 7. Low-penetration regime: Fairness (a) and CDF of the per-downloader throughput when $\hat{A} = 15$ (b), for different AP deployment strategies

AP deployment. Fig. 6 shows the average per-downloader throughput and delay for the different deployment strategies, as the number of active APs $\hat{A}$ varies. Overall, the performance at early deployment stages is satisfactory. The plots confirm that increasing $\hat{A}$ positively affects the downloading performance. However, it is also clear that the extension of the infrastructure deployment is more critical when the number of active APs is low. Indeed, activating 20 APs yields a 8-Mb/s throughput and 35-second delay gain over a 5-AP deployment, while the activation of additional 40 APs only leads to mere 3-Mb/s throughput and 15-second delay improvement.

The figure also highlights the impact of different AP deployment techniques in the low-penetration regime. As one can expect, the AP placement dictated by the Max-flow strategy guarantees the best performance in terms of both throughput and delay, while a Random deployment of APs yields the worst result. AP placements based on the Crowded and Contact approaches fall in between. If the performance ranking of the deployment strategies is constant throughout all values of $\hat{A}$, the same is not true for the relative gain. Indeed, when a few APs are activated, a well-planned deployment can result in a 200% throughput gain over a random placement. As the number of deployed APs grows, such an advantage is progressively reduced: in particular, when APs cover more than 50% of the road topology (i.e., $\hat{A} > 30$), using non-optimal approaches to AP deployment makes the performance quickly close in towards those achieved with a random placement. Finally, as the active APs tend to cover the whole region, optimal and non-optimal strategies yield similar performance.

Fairness. As shown in Fig. 7(a), the system favors downloaders travelling on the main roads when $\hat{A}$ is low, while user experience tends to be leveled as more and more APs are activated. The Max-flow deployment results in a slightly fairer

system, while no significant difference can be appreciated among non-optimal placement strategies.

The reason for the unfairness for small AP deployments is investigated in Fig. 7(b), for $\hat{A} = 15$. The plot reports the CDF of the per-downloader throughput, and shows a large heterogeneity in the amount of content obtained by different users. On the one hand, a significant percentage of downloaders, between 20% and 40% depending on the deployment strategy, experiences zero throughput. On the other, the luckiest 10% of downloaders enjoys a throughput between 16 and 20 Mb/s.

Although fairness in content downloading is not an objective of the max-flow problem[3], we point out that such unfairness is only marginally due to our formulation. Rather, it can be attributed to the very different conditions incurred in by downloaders in a realistic mobility trace, such as various traffic intensity on the roads they travel on, or different time intervals spent within coverage of the APs in their trip. This is confirmed by the limit curve in Fig. 7(b), which presents the CDF of the maximum achievable throughput, i.e., the throughput that each of the downloaders in the trace would experience if it were the only downloader in the network, with all resources and relays at its disposal. Not even in such ideal conditions one can guarantee fairness among all users, given their different trips. More pervasive AP coverages (Fig. 7(a)) or additional relaying opportunities (Fig. 5(c)) can help to reduce disparities, by providing transfer paths to downloaders travelling on secondary routes.

Transfer paradigm. The fraction of content downloaded through vehicular relays is shown in Fig. 8(a). Across almost

---

[3]We remark that, in an attempt to provide the downloaders with fairer performance, we have considered a max-min formulation instead of the max-flow one. However, due to the diversity in the downloader conditions highlighted next, no minimum positive throughput could be guaranteed.
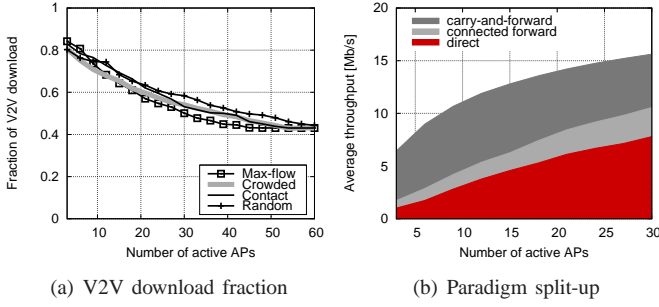
(a) V2V download fraction   (b) Paradigm split-up

Fig. 8. Low-penetration regime: Fraction of data downloaded through V2V (a) and transfer paradigm split-up of the per-downloader throughput (b). In the latter plot APs are deployed according to the Max-flow strategy
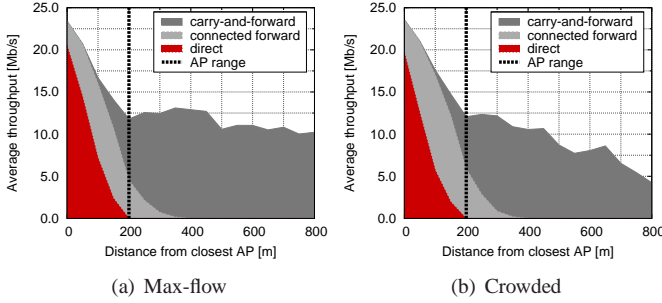

(a) Max-flow   (b) Crowded

Fig. 9. Low-penetration regime: Average per-downloader throughput as a function of the distance between the downloader and the closest AP

all AP deployment strategies, the V2V downloading fraction is around 0.8 when a low number of APs is activated, and then decreases as $\hat{A}$ grows, i.e., as a direct access to the infrastructure becomes more pervasive. This is a rather intuitive behavior that, however, yields an interesting result when coupled with the average per-downloader throughput, as in Fig. 8(b). There, we can observe that the fraction of content downloaded through V2V relaying is somewhat constant, contributing approximately 5 Mb/s to the overall throughput, regardless of the number of deployed APs. Also, the dominant relay paradigm is carry-and-forward.

A quite surprising result is that, even when APs fully cover the road topology (for $\hat{A} = 60$), V2V relaying is still widely employed in the downloading process. The reason behind such a phenomenon is unveiled in Fig. 9, portraying the average per-downloader throughput as a function of the downloader distance from the closest AP, for the Max-flow and Crowded (i.e., the most performing) strategies. The plot highlights the portions of traffic transferred through the different paradigms. As one would expect, direct I2V transfers can only occur within the AP transmission range, and a 2-hop connected forwarding reaches at most twice such a distance. Carry-and-forward is not distance-bounded, hence it can reach downloaders that are very far from APs.

However, a key observation is that V2V relaying frequently occurs within range of APs. In fact, at a distance of 100 m, i.e., half of the maximum transmission range of the AP, communication largely takes place through relays. The reason is that our model realistically accounts for the network-layer rate decrease with distance, hence making the use of high-rate multi-hop paths preferable to low-rate direct transfers. This explains why, even in presence of a pervasive AP coverage,

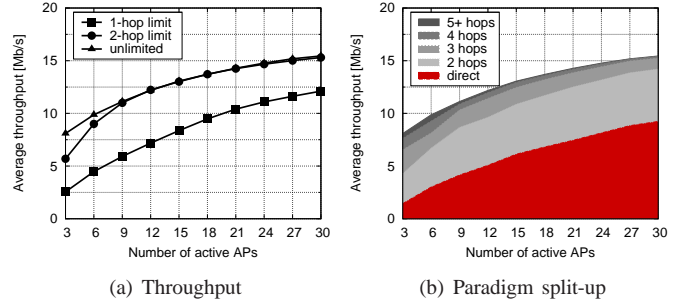(a) Throughput   (b) Paradigm split-up

Fig. 10. Low-penetration regime: Average per-downloader throughput (a) and transfer paradigm split-up (b), when APs are deployed according to the Max-flow strategy. In the latter plot, the number of relay hops in unconstrained

relaying is employed to improve the wireless resource utilization and, thus, the overall throughput.

Unlimited relaying. All previous results assumed a 2-hop limit in data transfers, basically constraining V2V relaying to one hop at most. We now relax this assumption and compare three different scenarios, where (i) only direct I2V communication is allowed, (ii) the 2-hop limit is enforced, and (iii) unlimited relaying is allowed.

Fig. 10(a) depicts the average throughput achieved in the three cases, when the APs are deployed according to the Max-flow strategy. It is clear that, in absence of relaying through vehicles, downloaders can only leverage direct contacts with the APs, which leads to a significantly lower throughput. Allowing a single relay between APs and downloaders yields a throughput gain between 150% and 20%, depending on the infrastructure coverage. Even more interestingly, considering transfers over 3-hop long or more yields almost no advantage over the case where a 2-hop limit is enforced.

In order to explain the latter effect, we fragment the downloading throughput measured in the former scenario, according to the number of hops traveled by packets to reach their destination. Fig. 10(b) shows how, even when unlimited hops are allowed, a large majority of the relayed data traffic arrives at destination in just two hops. There is a small probability of going through 3 hops, while 4 hops or more are almost never employed. When comparing Fig. 10(b) to Fig. 8(b), it is clear that the availability of additional hops, which grants more flexibility to the max-flow problem solution, only leads to minor adjustments that have a negligible impact on the overall downloading performance. Indeed, in the case of connected-forwarding transfers, we observed that, unless the APs are very sparse, the distance in hops between a downloader and the closest AP is rarely beyond 2, due to the network fleeting connectivity. In the case of carry-and-forward transfers, when the APs are not sparse, by the time a downloader can receive data that have traversed more than 2 hops, it is likely to be either under coverage of another AP or close enough to receive data in two hops from it.

Road environment. The average per-downloader throughput recorded in the three road topologies presented in Sec. VI is portrayed in Fig. 11(a), when the APs are deployed as dictated by the Max-flow strategy.

The overall performance trend in the new environments is the same as already observed in the urban scenario, thus our
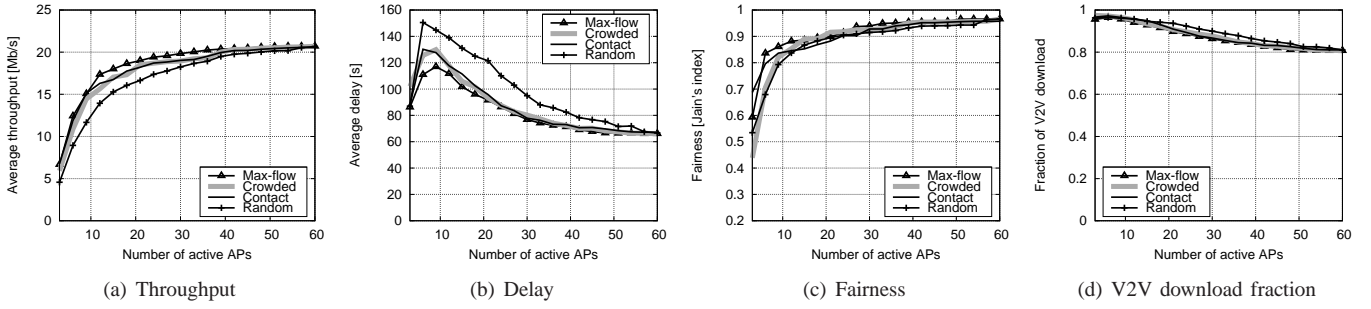
Fig. 12. High-penetration regime: Throughput (a), delay (b), fairness (c), and V2V downloading fraction (d) vs. $\hat{A}$, for different AP deployment strategies
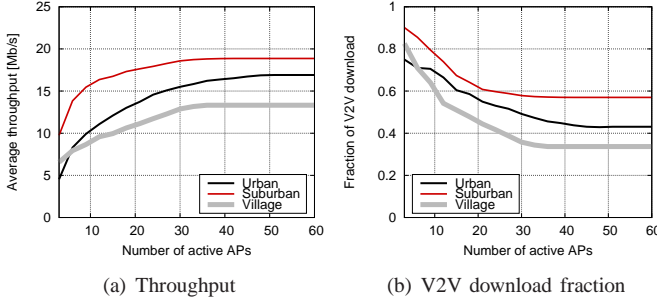

Fig. 11. Low-penetration regime: Average per-downloader throughput (a) and fraction of data downloaded through V2V (b) vs. $\hat{A}$, in different road environments. APs are deployed according to the Max-flow strategy

considerations on the impact of the AP deployment also hold for the village and suburban environments.

However, we can observe that the relative result in suburban and village environments differs from that measured in the urban case. On the one hand, the throughput in the village scenario is lower than in the urban one, with a significantly reduced utilization of V2V relaying. On the other, the suburban scenario yields higher V2V download fraction and per-downloader throughput.

The reason for these different behaviors is found in the diverse nature of vehicular traffic in the tree regions. By looking at Fig. 4(b), it is clear that fewer vehicles circulate in the village environment than in the urban one: thus, for a given p, fewer vehicles participate in the content downloading as relays. Moreover, the traffic is distributed over the road topology quite evenly, which makes it difficult to find an AP deployment that well covers most of the vehicular traffic. As a result, downloaders in the village scenario are penalized in terms of throughput.

In the suburban scenario, the car traffic volume is close to that observed in the urban environment, which means that the number of available relays in the two cases is similar. However, the suburban region is characterized by a few high-traffic thoroughfares and many low-traffic secondary roads. As the vehicular traffic is so concentrated, it is easier to deploy a few APs in the right locations; also, downloaders have higher chances of meeting many relays on their way. Thus, drivers in the suburban environment typically enjoy a higher throughput.

Summary. In the low-penetration regime, the early infrastructure deployment stages are critical. When just a few APs are activated, the policy chosen for their placement has a major impact on the user experience, as optimal deployments lead

to a throughput twice or three times higher than that observed with careless placements. Moreover, the activation of a few more (or less) APs dramatically affects the throughput, delay and fairness of the system. Since the downloading performance during early adoption phases will play an important role in attracting new users, the AP deployment should be carefully studied when introducing the technology. Conversely, the placement of too many APs may have a small impact on the downloading performance, while significantly increasing the deployment and maintenance costs.

One should not expect the system to be fair, or to have similar performance in different road environments, since the diversity in the routes traveled on by drivers lead to intrinsic differences in their download experience.

As a final remark, our results suggest that the complexity of designing multi-hop relaying protocols can be safely avoided, by limiting the process to one relay, without incurring in performance penalties. This confirms recent findings on bus networks [16], [22], which thus apply also to a more general vehicular downloading context.

C. High-penetration regime

In the high-penetration regime, we consider $p = 50\%$ and the max-flow problem is solved on the sampled DNTG. Once more, the default settings include the urban scenario, non-overlapping AP coverages, a 1% fraction of downloaders and a 2-hop limit in V2V relaying.

AP deployment. The overall performance is outlined in Fig. 12, for different extensions and strategies of the roadside infrastructure deployment. When comparing the results to those obtained in the low-penetration regime, we observe a significant improvement in the absolute value of the throughput, in Fig. 12(a), that now reaches more than 20 Mb/s – a clear effect of the increased availability of relays. The throughput growth is much faster as additional APs are deployed, with nearly optimal performance attained with as few as 15 active APs. This is due to the fact that relays can now easily compensate for undersized infrastructure, as also demonstrated by the extremely frequent utilization of V2V communication, in Fig. 12(d), employed in 80% to 98% of the transfers.

As far as delay is concerned, Fig. 12(b) exhibits a peculiar behavior. Given the high number of users, several downloaders happen to travel on secondary roads. For a very low number of APs, such roads are scarcely covered, hence a number of downloaders experience zero throughput. Their delay is not
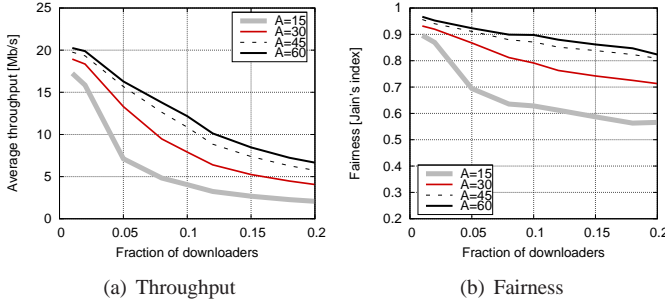
(a) Throughput   (b) Fairness

Fig. 13. High-penetration regime: Average per-downloader throughput (a) and fairness (b) vs. d, under the Max-flow strategy and for different AP deployment extensions
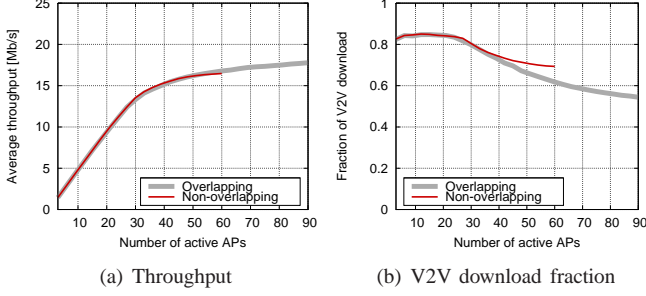

(a) Throughput   (b) V2V download fraction

Fig. 14. High-penetration regime: Throughput (a) and fraction of data downloaded through V2V (b) vs. $\hat{A}$, when $d = 0.05$. The case of overlapping and non-overlapping AP coverages are compared, under the Max-flow strategy

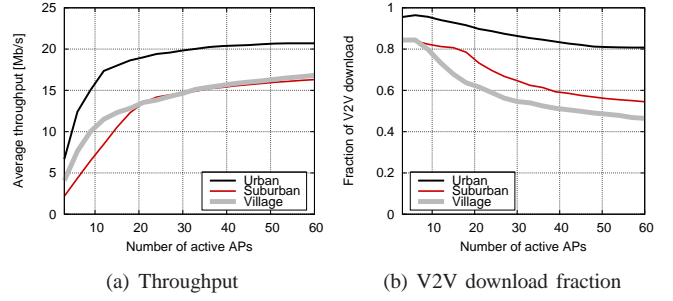(a) Throughput   (b) V2V download fraction

Fig. 15. High-penetration regime: Average per-downloader throughput (a), and fraction of data downloaded through V2V (b) vs. $\hat{A}$, in different road environments. APs are deployed according to the Max-flow strategy

accounted for, and the dominant contribution is limited to the few lucky fast downloaders. As the deployment becomes more widespread and $\hat{A}$ increases, more downloaders experience non-zero coverage time, including those on secondary roads where the chances to carry on the download are few and far between. For even denser deployments, such delays are mitigated by the availability of more APs.

Finally, the massive presence of relays helps to reduce the unfairness, in Fig. 12(c), as downloaders have high chances to meet relays, regardless of the route they take.

Concurrent downloaders. In presence of a wide diffusion of I2V and V2V communications, the downloading activity by users participating in the system is likely to grow. Thus, in the high-penetration regime, it is important to evaluate the impact of the amount of concurrent downloaders, i.e., users requesting some content during the same time interval.

Fig. 13 shows the impact of the concurrent downloader fraction d on the performance. We consider d ranging between 0.01 and 0.2, the latter representing a highly-loaded system, in which one out of five users is downloading data at any time instant. As one could expect, when the system load grows, increasing the number of APs comes in handy, and can noticeably improve throughput (Fig. 13(a)) and fairness (Fig. 13(b)). Also, increasing the demand (especially, for $d > 0.1$) reduces the per-user throughput, due to the augmented contention for the limited wireless resources. Less intuitively, fairness degrades as d grows: when the number of simultaneous downloaders increases, vehicles on secondary roads experience less channel contention, hence higher throughput than vehicles travelling on main (more crowded) roads.

Overlapping AP coverages. Given the beneficial effect of

additional APs when d is high, we study the impact of a further infrastructure extension by allowing AP coverages to overlap. Fig. 14 compares the throughput and V2V download fraction obtained when non-overlapping and overlapping AP coverages are allowed. The results have been obtained for a relatively high downloading demand, namely, $d = 0.05$. When overlapping among AP coverages is not allowed, only 60 candidate locations can be considered. Conversely, such a number grows to 90 when the coverage of any two candidate APs can overlap. Observe that, for a fixed number of APs, the possibility to have overlapping coverages leads to a marginal improvement in the per-downloader throughput. As shown by Fig. 14(b), the reason for this behavior is once more that the V2V traffic relaying tends to compensate for the lack of flexibility of the non-overlapping deployment.

Road environment. Fig. 15 shows the throughput and V2V download fraction in the urban, suburban and village scenarios. As already observed in the low-penetration regime, also in presence of high p the road topology has a major impact on the downloading process. However, the relative performance of the three scenarios are different with respect to those in Fig. 11. The highest throughput is now achieved in the urban scenario, while drivers in the suburban and village environments experience similarly worse performance.

The reason lies in the increased contention for resources, induced by the higher participation of vehicles in the network. In the urban scenario, many vehicles travel over different roads, which basically allows a spatial reuse of the wireless medium. The village scenario is similar to the urban one, in that vehicular traffic is quite evenly spread over the road topology; however, the lower number of vehicles reduces the availability of relays, as also evident from Fig. 15(b). In the suburban scenario, instead, a high vehicular density is concentrated on a few roads: the consequent channel congestion yields reduced per-downloader throughput.

Summary. The analysis in the high-penetration regime significantly differs from the early technology adoption phase. When the technology is mature and spread enough, the infrastructure deployment will play a minor role, and a few, randomly deployed APs will suffice to achieve near-optimal downloading performance. Indeed, V2V communication will be able to sustain the system, no matter the underlying AP placement. Pervasive non-overlapping APs will be needed only in case the technology attains a level of success such that

the number of concurrent downloaders grows well above the percentages today's recorded in wired networks. In this case, channel contention will become the primary constraint to the downloading performance, with I2V and V2V transfers contending for air time across the whole road topology. As such, redundant coverages will not yield a significant throughput gain and downloaders travelling on more crowded roads will experience worse performance.

VIII. IMPACT OF MAC AND PHYSICAL LAYER MODELING

The max-flow problem we formulate relies on a simplified model of channel access and RF signal propagation. Since our goal is to derive an upper bound to the performance achievable in a real-world deployment, these assumptions are not especially limiting. However, one may wonder about the impact that more realistic MAC and physical layer representations have on the system, i.e., how much their idealization contributes to shift the upper bound away from the actual performance. Including complex models of signal propagation and layer-2 protocols in the optimization problem is unfeasible, thus we rely on simulation to evaluate these aspects.

We consider the same reference scenario, road topology and vehicular mobility used for the urban environment in the previous section, with non-overlapping AP coverages and data transfers limited to 2 hops. Under these conditions, we assess the impact of our simplifying assumptions on the MAC and physical layers by employing ns-3. At the physical layer, the simulator computes the bit error rate accounting for the SINR and the signal modulation. We set the transmit output power to 16 dBm and the path loss exponent to 3; also, we included a log-normal shadowing with zero mean and standard deviation of 8 dB. At the MAC layer, we use IEEE 802.11a with the AARF rate adaptation algorithm [30]. We feed the optimal scheduling to the simulator, i.e., the scheduler provides (i) the amount of application-layer data that each AP has to send to relays or downloaders at a given frame, and (ii) the amount of application-layer data that every relay has to deliver to downloaders at a given frame. Vehicle association to APs is simulated, and vehicles periodically send a heartbeat message including their ID. Thus, an AP delivers the data to the intended relay or downloader after association; the AP also informs the relay about the identity of the downloader to which data have to delivered. Relays deliver the data to a downloader upon hearing the periodic heartbeat message.

Fig. 16 depicts the average throughput, delay, fairness and V2V downloading fraction, in the case of high-penetration regime. We report the results of the max-flow problem on the sampled DNTG and those obtained under the Crowded strategy. The optimization problem results are compared to ns-3 simulations. The qualitative behavior of the considered metrics derived through optimization and simulation closely match, and, even from the quantitative viewpoint, the difference is limited in all cases. This demonstrates that it is the optimality of the scheduling that plays the most important role in determining the downloading experience of the users, rather than the local channel access coordination of individual transmissions or the propagation conditions model.

Similar results (included in Appendix A-4) have been obtained for the less critical case of the low-penetration regime.

We stress that plots are limited to 30 APs along the x-axis, since larger simulations would have required an exceedingly long computational time. This further underscores the usefulness of our formulation, which makes much more extensive evaluations of the downloading system feasible.

IX. CONCLUSIONS

We proposed a framework based on time-expanded graphs for the study of content downloading in vehicular networks. Our approach allows to capture the space and time network dynamics, and to formulate a max-flow problem whose solution provides an upper bound to the system performance. Through a graph-sampling technique, we solved the problem for realistic, large-scale traces. Simulation results showed that the physical- and MAC-layer assumptions on which the framework relies have a minor impact, leading to a tight upper bound.

The major findings in our analysis are as follows.
(i) Two regimes, characterized by different performance and impact of the system settings, emerge at different technology penetration rates. In a urban scenario, the watershed arises when 20-30% of the vehicles participate in the network.
(ii) The strategy and the extension of the AP deployment play a major role in the low-penetration regime, with well-planned deployments leading to a throughput twice or three times higher than that observed under a careless placement. In the high-penetration regime, instead, even a random AP deployment works well and the pervasiveness of the APs becomes important only for high downloading demand.
(iii) The contribution to performance of V2V traffic relaying is critical. It can compensate for reduced coverage as well as for a non-optimal AP placement, with such an effect becoming more and more evident as the technology penetration rate increases. The contribution of V2V communications remains relevant even under a pervasive AP deployment and in both penetration regimes, as optimal scheduling tends to favor high-rate V2V transfers over low-rate I2V communications.
(iv) Knowledge of user mobility is paramount to the system performance, since most of the V2V traffic relaying takes place through the carry-and-forward paradigm. However, the complexity of multi-hop protocols can be limited to one relay, as the contribution of transfers over a higher number of hops is negligible. An interesting direction for future research is therefore the design of protocols that let the roadside infrastructure acquire accurate estimates of the vehicles encounter opportunities, and the definition of a scheduling algorithm that effectively leverages such information. We remark that, by using edges with probabilistic weights, our graph-based model could account for the uncertainty in the mobility estimates.
(v) The structure of the road topology and the route followed by vehicles determine the downloading performance experienced by the users. Thus, one should adapt the system configuration to the characteristics of the road environment. In any case, some unfairness should be expected unless there is a pervasive presence of APs and relays, and the number of downloaders is not overwhelming.
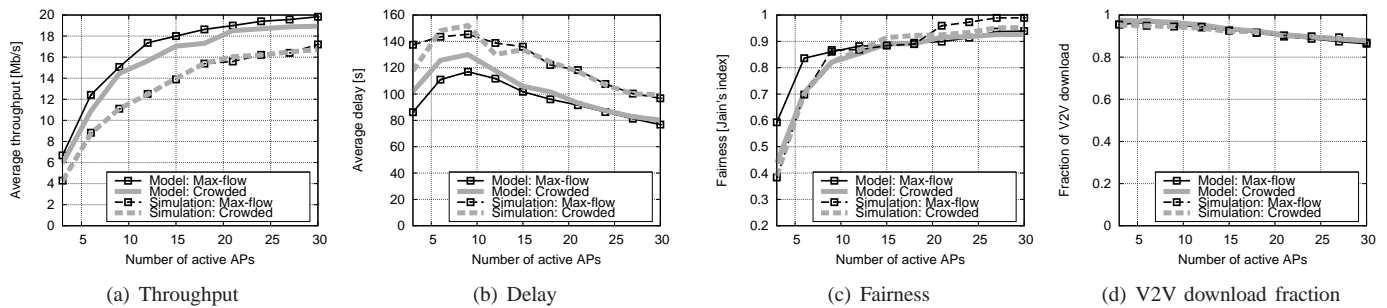
Fig. 16. High-penetration regime: Average per-downloader throughput (a), delay (b), fairness (c) and V2V downloading fraction (d), as the number of APs and the deployment strategy vary. The results obtained through model and simulation are compared

REFERENCES

[1] B.S. Arnaud, "iPhone slowing down the Internet – Desperate need for 5G R&E networks," May 2010, http://billstarnaud.blogspot.com/2010/04/iphone-slowing-down-internet-desperate.html.

[2] U. Paul, A.P. Subramanian, M.M. Buddhikot, S.R. Das, "Understanding traffic dynamics in cellular data networks," *IEEE INFOCOM,* Shanghai, China, Apr. 2011.

[3] Z. Zheng, P. Sinha, S. Kumar, "Alpha coverage: bounding the interconnection gap for vehicular Internet access," *IEEE INFOCOM,* Rio de Janeiro, Brazil, Apr. 2009.

[4] Z. Zheng, Z. Lu, P. Sinha, S. Kumar, "Maximizing the contact opportunity for vehicular Internet access," *IEEE INFOCOM,* San Diego, CA, Mar. 2010.

[5] M. Fiore, J. M. Barcelo-Ordinas, "Cooperative download in urban vehicular networks," *IEEE MASS*, Macau, China, Oct. 2009.

[6] D. Hadaller, S. Keshav, T. Brecht, S. Agarwal, "Vehicular opportunistic communication under the microscope," *ACM MobySys,* San Juan, Puerto Rico, June 2007.

[7] B. B. Chen, M. C. Chan, "MobTorrent: A framework for mobile Internet access from vehicles," *IEEE INFOCOM,* Rio de Janeiro, Brasil, Apr. 2009.

[8] J. Zhao, T. Arnold, Y. Zhang, G. Cao "Extending drive-thru data access by vehicle-to-vehicle relay," *ACM VANET,* San Francisco, CA, Sept. 2008.

[9] J. Ott, D. Kutscher, "Drive-thru Internet: IEEE 802.11b for "Automobile" users," *IEEE INFOCOM*, Hong Kong, China, Mar. 2004.

[10] V. Bychkovsky, B. Hull, A. K. Miu, H. Balakrishnan, S. Madden, "A measurement study of vehicular Internet access using in situ Wi-Fi networks," *ACM MobiCom,* Los Angeles, CA, Sept. 2006.

[11] G. Marfia, G. Pau, E. Giordano, E. De Sena, M. Gerla, "Evaluating vehicle network strategies for downtown Portland: opportunistic infrastructure and importance of realistic mobility models," *ACM MobiOpp,* San Juan, Puerto Rico, June 2007.

[12] Y. Ding, C. Wang, L. Xiao, "A static-node assisted adaptive routing protocol in vehicular networks," *ACM VANET,* Montreal, Canada, Sept. 2007.

[13] C. Lochert, B. Scheuermann, C. Wewetzer, A. Luebke, M. Mauve, "Data aggregation and roadside unit placement for a VANET traffic information system," *ACM VANET,* San Francisco, CA, Sept. 2008.

[14] A. Abdrabou, W. Zhuang, "Probabilistic delay control and road side unit placement for vehicular ad hoc networks with disrupted connectivity," *IEEE JSAC*, vol. 29, no. 1, Jan. 2011.

[15] V. Kone, H. Zheng, A. Rowstron, B.Y. Zhao, "On infostation density of vehicular networks," *WICON,* Singapore, Mar. 2010.

[16] A. Balasubramanian, B.N. Levine, A. Venkataramani, "Enhancing interactive Web applications in hybrid networks," *ACM/IEEE MobiCom*, San Francisco, CA, Sept. 2008.

[17] S. Yoon, D. T. Ha, H. Q. Ngo, C. Qiao, "MoPADS: A mobility profile aided file downloading service in vehicular networks," *IEEE Trans. on Veh. Tech.*, vol. 58, no. 9, Nov. 2009.

[18] A. Nandan, S. Das, G. Pau, M. Gerla, M. Y. Sanadidi, "Co-operative downloading in vehicular ad-hoc wireless networks," *IEEE WONS,* St.-Moritz, Switzerland, Jan. 2005.

[19] M. Gerla, M. Gruteser, "Vehicular networks: Applications, protocols, and testbeds," in *Emerging Wireless Technologies and the Future Mobile Internet,* Eds.: D. Raychaudhuri, M. Gerla, May 2011.

[20] M. Sardari, F. Hendessi, F. Fekri, "Infocast: A new paradigm for collaborative content distribution from roadside units to vehicular networks," *IEEE SECON,* Rome, Italy, June 2009.

[21] O. Trullols-Cruces, J. Morillo, J. M. Barcelo-Ordinas, J. Garcia-Vidal, "A cooperative vehicular network framework," *IEEE ICC,* Dresden, Germany, June 2009.

[22] N. Banerjee, M.D. Corner, D. Towsley, B.N. Levine, "Relays, meshes, base stations: Enhancing mobile networks with infrastructure," *ACM/IEEE MobiCom,* San Francisco, CA, Sept. 2008.

[23] D. Hay, P. Giaccone, "Optimal routing and scheduling for deterministic delay tolerant networks," *IEEE WONS,* Snowbird, UT, Feb. 2009.

[24] F. Malandrino, C. Casetti, C.-F. Chiasserini, M. Fiore, "Content downloading in vehicular networks: what really matters," *IEEE INFOCOM Mini-Conference*, Shanghai, China, Apr. 2011.

[25] G. Iosifidis, I. Koutsopoulos, G. Smaragdakis, "The impact of storage capacity on end-to-end delay in time varying networks," *IEEE INFOCOM*, Shanghai, China, Apr. 2011.

[26] L. R. Ford, D. R. Fulkerson, *Flows in networks*, Princeton University Press, Princeton, NJ, 1962.

[27] L. Lovasz, "Random walks on graphs. a survey," *Combinatorics*, 1993.

[28] G. Hertkorn, P. Wagner, "The application of microscopic activity based travel demand modelling in large scale simulations," *World Conference on Transport Research*, Istanbul, Turkey, 2004.

[29] F. Aidouni, M. Latapy, C. Magnien, "Ten weeks in the life of an eDonkey server," *International Workshop on Hot Topics in Peer-to-Peer Systems (Hot-P2P)*, Rome, Italy, 2009.

[30] M. Lacage, M. Hossein Manshaei, T. Turletti, "IEEE 802.11 rate adaptation: a practical approach," *ACM MSWiM*, Venice, Italy, Oct. 2004.

Francesco Malandrino (S'10) graduated (summa cum laude) in Computer Engineering from Politecnico di Torino in 2008. He received his PhD from Politecnico di Torino in 2011. From 2010 till 2011, he has been a visiting researcher at the University of California at Irvine. His interests focus on wireless and vehicular networks and infrastructure management.

Claudio Casetti (M'05) graduated in Electrical Engineering from Politecnico di Torino in 1992 and received his PhD in Electronic Engineering from the same institution in 1997. He is an Assistant Professor at the Dipartimento di Elettronica of Politecnico di Torino. He has coauthored more than 130 journal and conference papers in the fields of networking and holds three patents. His interests focus on ad hoc wireless networks and vehicular networks.

Carla-Fabiana Chiasserini (M'98, SM'09) received her Ph.D. from Politecnico di Torino in 2000. She has worked as a visiting researcher at UCSD in 1998–2003, and she is currently an Associate Professor at Politecnico di Torino. Her research interests include architectures, protocols, and performance analysis of wireless networks. Dr. Chiasserini has published over 190 papers in prestigious journals and leading international conferences, and she serves as Associated Editor of several journals.

Marco Fiore (M'05) is an Assistant Professor at INSA Lyon, and an INRIA researcher within the UrbaNet team hosted by the CITI Lab. He received M.Sc degrees from the University of Illinois at Chicago and Politecnico di Torino, in 2003 and 2004, respectively, and his PhD from Politecnico di Torino, in 2008. He has been a visiting researcher at Rice University and Universitat Politecnica de Catalunya. His research interests are in the field of mobile and vehicular networking.