Content Downloading in Vehicular Networks: What Really Matters

Francesco Malandrino
Dipartimento di Elettronica, Politecnico di Torino
malandrino@tlc.polito.it

Claudio Casetti
Dipartimento di Elettronica, Politecnico di Torino
casetti@polito.it

Carla-Fabiana Chiasserini
Dipartimento di Elettronica, Politecnico di Torino
chiasserini@polito.it

Marco Fiore
Université de Lyon, INRIA
INSA-Lyon, CITI Lab
marco.fiore@insa-lyon.fr

Abstract—Content downloading in vehicular networks is a topic of increasing interest: services based upon it are expected to be hugely popular and investments are planned for wireless roadside infrastructure to support it. We focus on a content downloading system leveraging both infrastructure-to-vehicle and vehicle-to-vehicle communication. With the goal to maximize the system throughput, we formulate a max-flow problem that accounts for several practical aspects, including channel contention and the data transfer paradigm. Through our study, we identify the factors that have the largest impact on the performance and derive guidelines for the design of the vehicular network and of the roadside infrastructure supporting it.

I. INTRODUCTION

Within the next few years, the proliferation of in-vehicle communication interfaces is envisaged to become a reality that will enable new Intelligent Transportation System applications. Beside critical safety services, content downloading is expected to be widely popular with users of a vehicular network. Examples abound, such as drivers interested in downloading enhanced local maps, possibly including current traffic conditions, or passengers wishing to download media-rich data files and touristic information from the Internet.

As a result, content downloading in vehicular networks has received increasing attention from the research community. On the one hand, the availability of Infrastructure-to-Vehicle (I2V) communication capabilities, based on high-throughput Dedicated Short-Range Communication (DSRC) technologies, is seen as an opportunity for bulk transfers to mobile nodes that would not be otherwise possible or scalable through the existing 2G/3G infrastructure. On the other hand, the introduction of Vehicle-to-Vehicle (V2V) connectivity has fostered a number of proposals to exploit the cooperation among vehicular users so as to improve their downloading performance. In particular, V2V-based approaches are especially attractive when one considers that the infrastructure coverage will be spotty at initial stages, and hardly seamless even at later ones.

Previous works on content downloading in vehicular networks have dealt with individual aspects of the process, such as the deployment of roadside APs [1]–[3], the performance evaluation of I2V communication [4], the network connectivity [5], [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 order to fill such a gap, we pose the following questions: (i) which is the maximum downloading performance theoretically achievable through DSRC-based I2V/V2V communication, in a given mobility scenario? (ii) what are the factors that mainly determine such a performance?

To answer these questions, 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 attainable in practice.

Although the problem formulation and the performance we derive are interesting per se, we also exploit our optimal solution to benchmark several AP deployment strategies, and to obtain useful hints for a practical implementation of content downloading relying on I2V and V2V communication.

Our framework employs a DTN time-invariant graph, similar to that in [9]. Unlike [9], however, we do not assume the contacts between mobile nodes to be atomic but allow them to have arbitrary duration, and we build the network graph so as to account for the presence of roadside infrastructure and channel contention. Such an approach allows us to significantly enhance the AP deployment for cooperative vehicular downloading proposed in [3], since we maximize the actual throughput instead of a metric, provide the optimal solution instead of an approximation, and model previously neglected channel contention and data rate adaptation.

II. 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 may be 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 consider the following data transfer paradigms:

- **direct** transfers, 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 multihop 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 such downloader sooner.

Note that the union of these paradigms covers the entire set of possibilities for data transfer through 12V/V2V communication. We stress that connected forwarding and carry-and-forward are inherently multi-hop paradigms, in presence of which we assume that downloaders are selfish, i.e., they never act as relays. Furthermore, since we are interested in deriving an upper bound to the system performance, we do not address data traffic scheduling or relay selection, but assume the availability of preemptive knowledge of vehicular trajectories and perfect scheduling of data transmissions.

From the viewpoint of the network system, we assume that any 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 if their distance is below or equal to their maximum radio range, which, without loss of generality, we assume to be common to all network nodes. We consider that the network nodes operate on the same frequency channel. The nodes share the channel bandwidth allocated for service applications using an IEEE 802.11-based MAC protocol, with RTS/CTS handshake.

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 graph that represents the temporal network evolution (Sec. III). By using this graph, we formulate a max-flow problem whose solution matches our goals (Sec. IV).

### III. Dynamic Network Topology Graph

We generate a dynamic network topology graph (DNTG) from a vehicular mobility trace, considering that on the corresponding road layout there are: (i) a set of $A$ candidate locations ($a_i$, $i = 1, \ldots, A$) where APs could be located, (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 through 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:

1. (i) the quality level of the link between the two nodes; specifically, we take as link quality metric the achievable data rate at the network layer, which depends on the distance between the two nodes (other metrics, such as the transmission data rate on the wireless channel, could be considered as well);
2. (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;
3. (iii) a 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 of the real-world communication, such as channel contention and the presence of hidden nodes.

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

Within each frame $k$, a directed edge $(v^k_i, v^k_j)$ exists from vertex $v^k_i \in \mathcal{R}^k$ to vertex $v^k_j \in V^k$ 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 \( w(v^k_i, v^k_j) \), equal to the rate of that contact event. The set including such edges is defined as \( L^k_a \). Similarly, a directed edge \((a^k_i, v^k_j)\) exists from vertex \( a^k_i \in A^k \) to vertex \( v^k_j \in V^k \) if a contact between the candidate AP \( a_i \) and the vehicle \( v_j \) is active during frame \( k \). Again, these edges are associated with weights \( w(a^k_i, v^k_j) \), corresponding to the contact event rate, and their set is defined as \( L^k_a \).

A directed edge \((v^k_i, v^k_{i+1})\) is also drawn from any vertex \( v^k_i \in R^k \) to any vertex \( v^k_{i+1} \in R^{k+1}, 1 \leq k < F \). While the edges in \( L^k_a \) and \( L^k_a \) represent transmission opportunities, those of the form \((v^k_i, v^k_{i+1})\) model the possibility that a non-downloader vehicle \( v_i \) physically carries some data during its movement 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 of such edges to take on an infinite value. A directed edge \((a^k_i, a^k_{i+1})\) of infinite weight is also drawn between any vertices representing the same candidate AP at two consecutive frames, i.e., from \( a^k_i \in A^k \) to \( a^k_{i+1} \in A^{k+1} (1 \leq k < F) \). We refer to the edges of the kind \((v^k_i, v^k_{i+1})\) or \((a^k_i, a^k_{i+1})\) as intra-nodal.

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

The DNTG is therefore a weighted directed graph, representing the network topology evolution over time. An example of DNTG is given in Fig. 1, in presence of one AP and three vehicles \( v_1, v_2, v_3 \), with \( v_1 \) being a downloader and \( v_2, v_3 \) possibly acting as relays. There, contact events separate different frames, that correspond to rows of vertices in the DNTG, where intra-nodal edges connect vertices representing the same vehicle or candidate AP over time. To limit the graph size, in this example we assume the achievable network-layer rate \( w \) to be constant during the entire lifetime of a link; in our performance evaluation, instead, we consider a more complex model, which accounts for realistic variations of the rate as a function of the distance between the nodes. Also, note that the graph allows the capture of 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 vehicle \( v_3 \), as path \( A \).

IV. 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 \( \alpha \) to \( \omega \), 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:

\[
\max \sum_{k=1}^{F} \sum_{v^k_i \in D^k} x(v^k_i, \omega)
\]

The max-flow problem needs to be solved taking into account several constraints due to, e.g., flow conservation, maximum number of APs that can be activated, and channel access. We detail such constraints below.

A. Constraints

Non-negative flow and flow conservation: the flow on every existing edge must be greater than or equal to zero. Also, for any vertex in the DNTG, the amount of flow entering the vertex must equal the amount of outgoing flow.

Channel access: since we consider an IEEE 802.11-based MAC scheme with RTS/CTS and we assume unicast transmissions, two or more of the following events cannot take place simultaneously for a tagged vehicle, 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;
6) a neighboring AP transmits to any vehicle.

We point out that we do not model the fine scheduling of packets transmitted within each frame. Rather, we only consider the total amount of data carried by each flow. Also, in a neighboring vehicle receiving data is accounted for, due to the use of RTS/CTS. Considering that: 1) is a subcase of 2); 3) is a subcase of 4); 5) is a subcase of 6), for the generic vertex \( v^k_i \in V^k \) and for any frame \( k \), we have:

\[
\sum_{v^k_i \in R^k \land v^k_j \in V^k} \frac{x(v^k_i, v^k_j)}{w(v^k_i, v^k_j)} + \sum_{v^k_i \in R^k \land v^k_j \in V^k} \frac{x(v^k_i, v^k_j)}{w(v^k_i, v^k_j)} + \sum_{a^k_i \in A^k \land v^k_j \in V^k} \frac{x(a^k_i, v^k_j)}{w(a^k_i, v^k_j)} \leq \tau^k
\]

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

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 \( a^k_i \in A^k \), we have:

\[
\sum_{v^k_i \in V^k} \frac{x(a^k_i, v^k_j)}{w(a^k_i, v^k_j)} \leq \tau^k
\]
with a communication interface and is willing to participate in 92 candidate locations, shown in Fig. 2(b).

between two adjacent APs is at least equal to 150 m, resulting in possible locations along the roads such that the distance vehicles participating in the network.

concurrently request content is assumed to be 1% of the max-flow problem in Sec. IV provides the AP deployment solution.

tor [10]. In Fig. 2(a), we portray the road layout, highlighting ETH Zurich, through a multi-agent microscopic traffic simulator [2].

We consider a conservative technology penetration rate, i.e., we assume a penetration rate of 10%. In Fig. 2(a), we portray the road layout, highlighting the different traffic volumes observed over each road segment.

We consider a conservative technology penetration rate, i.e., that only a fraction of the vehicles, namely 10%, is equipped with a communication interface and is willing to participate in the content downloading process, either as relays or as downloaders. Also, the number of mobile downloaders that concurrently request content is assumed to be 1% of the vehicles participating in the network.

AP locations are selected picking all intersections and the possible locations along the roads such that the distance between two adjacent APs is at least equal to 150 m, resulting in 92 candidate locations, shown in Fig. 2(b).

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 [4].

We limit the maximum node transmission range to 200 m, since, as stated in [4], this distance allows the establishment of a reliable communication in 80% of the cases.

Given that \( \hat{A} \) locations have to be activated, the solution of the max-flow problem in Sec. IV provides the AP deployment that maximizes the aggregate download throughput. We benchmark the performance of our optimal strategy (hereinafter referred to as Max-flow strategy) against the following AP deployment policies:

**Random:** \( \hat{A} \) locations are randomly selected among the candidate ones, according to a uniform distribution;

**Crowded:** it picks the \( \hat{A} \) locations whose coverage area exhibits, over time, the highest vehicular density;

**Contact:** it selects the \( \hat{A} \) locations that maximize the sum of the contact opportunities between vehicles and APs [2]. Specifically, for each vehicle, the contact opportunity is expressed as the fraction of the road segment lengths traveled while under coverage of at least one AP.

Once the active AP locations are determined according to each of the above three strategies, they are used in the max-flow problem formulation to fix the values of the binary variables \( y_i \). Since the system throughput is obtained as the solution of the max-flow problem given the selected AP locations \( y_i \), the results we show represent the best performance one can achieve with each deployment strategy.

Fig. 3 shows the average per-downloader throughput for different deployment strategies, as a function of the number of active APs \( \hat{A} \). In particular, Fig. 3(a) portrays the performance of the Max-flow, Random, Crowded and Contact placement policies when only the direct transfer paradigm is allowed. It is clear that, in absence of relaying through vehicles, a random deployment of APs results in a throughput that is typically well below 50% of that attained under the optimal AP placement. Interestingly, the Crowded and Contact strategies yield a very similar performance, at around 80% of the optimum.

When the maximum number of hops the data are allowed to go through is increased to two, in Fig. 3(b), or it is unlimited, in Fig. 3(c), we observe that the relative performance of the strategies in presence of V2V communication does not change, with the Random policy always resulting in the lowest average throughput, and the Crowded and Contact policies performing similarly and better than the former. However, when we relax the limit in the number of hops, the performance achieved under the different deployment strategies tends to close in on the optimal throughput achieved by the Max-flow placement policy. This is especially true when the number of APs is small, and for the Random placement policy in particular: indeed, increasing the opportunity for traffic relaying mitigates the sub-optimality of the AP deployment that emerged under the direct transfer case.

At last, we point out that we observed acceptable latencies under all strategies. E.g., with \( \hat{A} = 15 \), 35% of data are delivered to downloaders in less than 10 s, 50% in less than 40 s, while 80% experiences a delay below 100 s.

Based on the above results, we conclude that a simple Crowded strategy can match the performance of a more complex policy such as the Contact strategy, and it can achieve a performance that is consistently within 80% of the optimum.

Looking again at Fig. 3, we now comment on the impact of the multi-hop paradigms, i.e., connected forwarding and carry-and-forward. By comparing Figs. 3(a)–3(c), it is clear that V2V communication can greatly enhance the performance.
that the amount of data downloaded via one relay remains clearly more frequent. However, it is interesting to observe that the presence of APs becomes more pervasive, direct transfers are less prevalent when the number of deployed APs is small. As the number of APs increases, direct transfers and reduces the airtime for multi-hop ones. In Fig. 4(a), the hop limit is set to 2, thus the plot portrays which portion of the average per-downloader throughput is due to direct transfers and which is instead attained using one relay: the latter largely dominates the former when the number of deployed APs is small. As the presence of APs becomes more pervasive, direct transfers are clearly more frequent. However, it is interesting to observe that the proportion of throughput achieved through direct and multi-hop transfers does not change when the limit on the number of allowed hops is removed, in Fig. 4(b). There, we can also note the small contribution due to transfers over 3 or more hops, especially for $A \geq 10$. As a conclusion, the comparison between Fig. 4(a) and Fig. 4(b) suggests that the complexity due to the use of more than one relay can be avoided without significant harm.

To summarize, we draw the following conclusions:
- traffic relaying, through either connected forwarding or carry-and-forward, can significantly increase the average per-downloader throughput, even when the road layout is largely covered by APs;
- multi-hop transfers involving more than one relay are scarcely beneficial to the downloading process.

VI. CONCLUSION

We investigated the main factors affecting the performance of content downloading in vehicular networks, by formulating and solving a max-flow problem over a graph representing a realistic vehicular trace. Our major findings are that a density-based AP deployment yields performance close to the optimum, and that multi-hop traffic delivery is beneficial, although the gain is negligible beyond 2 hops from the AP.

VII. ACKNOWLEDGMENT

This work was supported by the European Union through the FP7 NoE EuroNF.

REFERENCES


