



**HAL**  
open science

## Vehicles as Big Data Carriers: Road Map Space Reduction and Efficient Data Assignment

Benjamin Baron, Prométhée Spathis, Hervé Rivano, Marcelo Dias de Amorim

► **To cite this version:**

Benjamin Baron, Prométhée Spathis, Hervé Rivano, Marcelo Dias de Amorim. Vehicles as Big Data Carriers: Road Map Space Reduction and Efficient Data Assignment. VTC2014-Fall - IEEE 80th Vehicular Technology Conference, Sep 2014, Vancouver, Canada. pp.1-5, 10.1109/VTC-Fall.2014.6966227 . hal-00994848

**HAL Id: hal-00994848**

**<https://hal.sorbonne-universite.fr/hal-00994848>**

Submitted on 22 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Vehicles as big data carriers: Road map space reduction and efficient data assignment

Benjamin Baron<sup>◊</sup>, Prom  th  e Spathis<sup>◊</sup>, Herv   Rivano<sup>\*</sup>, and Marcelo Dias de Amorim<sup>◊,†</sup>

<sup>◊</sup>UPMC Sorbonne Universit  s    <sup>\*</sup>Inria, U. de Lyon, INSA-Lyon CITI    <sup>†</sup>CNRS

**Abstract**—We advocate the use of a data shuttle service model to offload bulk transfers of delay-tolerant data from the Internet onto standard vehicles equipped with data storage capabilities. We first propose an embedding algorithm that computes an offloading overlay on top of the road infrastructure. The goal is to simplify the representation of the road infrastructure as raw maps are too complex to handle. In this overlay, each logical link maps multiple stretches of road from the underlying road infrastructure. We formulate then the data transfer assignment problem as a novel linear programming model that determines the most appropriate logical paths in the offloading overlay for a data transfer request. We evaluate our proposal using actual road traffic counts in France. Numerical results show that we can satisfy weekly aggregate requests in the petabyte range while achieving cumulative bandwidth above 10 Gbps with a market share of 20% and only one terabyte of storage per vehicle.

## I. INTRODUCTION

We propose a disruptive offloading service model to handle the growth of demand for bandwidth-intensive background services that have strained the Internet infrastructure to its limits (such as maintenance activities, data migration, and online backups) [1]. Our service exploits the delay-tolerant nature of the background Internet traffic to *deliver massive amount of data over the conventional road infrastructure*, relieving the Internet from this traffic while providing a reliable and robust transfer. The idea is to take advantage of the increasing number of journeys involving vehicles, which remain by far the most commonly used mode of transport. We argue in favor of equipping standard electric vehicles with data storage capability as part of the recent development in the automotive sector. Vehicles so equipped act as data shuttles delivering data in large quantities from the producer repositories to the destination points. We consider the model in which electric vehicles charge or replace their depleted batteries at charging or swapping stations, respectively. In our model, we propose that these stations serve as *offloading spots* where vehicles can load/deliver data (making all the process totally transparent to the driver). Thus, vehicles take turns carrying the data to its destination by loading and unloading it at offloading spots.

Related proposals suggest offloading Internet data onto other transmission media such as planes [2], postal services [3], or even birds [4]. These approaches allow saving some bandwidth from the Internet and even reducing the costs compared to other solutions such as renting a dedicated line. Nevertheless, they fail to achieve massive offloading and can be used only in very specific situations. As we will show later in this paper, our solution is more ambitious and can achieve rates of several

gigabytes per second for a global amount of data reaching a petabyte per week.

A major challenge toward our vision is *how to assign a data transfer request onto the road infrastructure*. We provide a mathematical optimization formulation of this problem. We first propose a mapping algorithm that creates an offloading overlay network on top of the road infrastructure. Nodes in the overlay network are connected through virtual links, which correspond to logical paths connected by multiple stretches of road in the underlying road infrastructure. Each virtual link is characterized by a set of attributes such as delay, capacity, and data losses. The offloading overlay network provides an abstract view of the resources that may be allocated to a particular data transfer. Second, we formulate the data transfer assignment problem as a novel linear programming model that determines the optimal logical paths consisting of a set of offloading spots connected together by logical links in the offloading overlay.

We provide numerical results using actual road traffic counts in France. The offloading spots are determined to ensure full coverage of all data transfer requests. The search space is further reduced by restricting the physical paths that satisfy the delay requirements.

In summary, our main contributions in this paper are:

- **Space reduction.** We design a mapping algorithm that creates the offloading overlay that mitigates the complexity of the road infrastructure and gives a coherent representation of the underlying resources.
- **Data transfers.** We formulate the data transfer assignment problem as a novel linear programming model. The model provides the optimal logical paths on the offloading overlay for a given data transfer request.
- **System capacity.** We evaluate the offloading service model for the data transfer assignment problem using actual road traffic counts of the main roads of France. The results show impressive numbers, with potential transfers in the order of a petabyte during a one-week time window.

## II. PRINCIPLES

We consider a bimodal data transportation service that combines the legacy Internet with flows of vehicles traveling roads. Our offloading service aims at overcoming the limitations of the Internet in terms of capacity when it comes to transfer large amounts of delay-tolerant data. We assume that vehicles are equipped with some storage capacity and a network of offloading spots provides the locations where vehicles can

load/deliver data. Offloading spots could be battery charging and swapping stations as the ones Tesla is currently rolling out both in Europe and North America [5]. At offloading spots, data can be either transferred onto the storage devices of the vehicles or the latter can be exchanged for preloaded storage devices that match the destination of the vehicle.<sup>1</sup>

This offloading service allows to relieve the Internet of the background traffic by transferring it on the road infrastructure using vehicles that will take turns to carry the data to its destination. Hence, the background traffic will not hinder both the elastic and interactive traffic, which are under stricter constraints (in terms of delay and throughput) [6]. The service further provides a robust transfer of high volumes of data, as the transfer is spread over multiple independent vehicles going in different directions. Finally, we assume the on-board storage cannot be accessed by the drivers and the data is encrypted, preventing anyone from reading the content.

### III. OFFLOADING SERVICE MODEL AND NOTATIONS

The problem we introduce is twofold: (1) to propose a mapping algorithm to create the offloading overlay network on top of the road infrastructure and (2) to model the data transfer assignment problem with a linear programming model that aims at maximizing the revenue on the offloading overlay.

#### A. Offloading overlay

The road infrastructure is represented as a directed graph  $G^R = (N^R, L^R)$  where  $N^R$  and  $L^R$  refer to the set of nodes (junctions and stations where vehicles can charge/replace their depleted batteries) and links (stretches of road), respectively.

The road infrastructure has many characteristics to be taken into account, such as the flow of vehicles from an origin to a destination, the capacity, the free-flow speed, and the travel time of the stretches of road. We introduce an *offloading overlay* to aggregate these properties in a simplified representation of the road infrastructure. The offloading overlay is a bijection from the stations of the road infrastructure to *offloading spots* and a mapping from physical paths between pairs of stations to *logical links* connecting pairs of offloading spots.

We denote the offloading overlay by a directed graph  $G^O = (N^O, L^O)$ , where  $N^O$  and  $L^O$  refer to the set of nodes (offloading spots mapping the stations) and links (logical links between the offloading spots), respectively. The logical links are characterized a logical by the weighted travel time  $t(i, j)$ , aggregated capacity  $c(i, j)$ , and data leakage  $\mathcal{L}(i, j)$ , that is the proportion of data lost on logical link  $(i, j) \in L^O$ . In Fig. 1, we show an example of a possible realization of an overlay on top of the French road infrastructure.

#### B. Data transfer assignment problem

We consider data transfer requests, denoted  $\mathcal{R}^{st}$ , where data has to be delivered from a source offloading spot  $s \in N^O$  to a target offloading spot  $t \in N^O$ , within a maximal delay  $\mathcal{T}^{st}$ , and with a leakage tolerance  $\mathcal{L}^{st}$ . We denote by  $\mathcal{R}$  the matrix

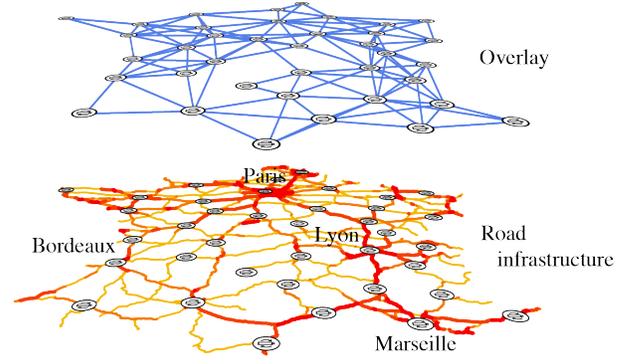


Fig. 1: France main roads infrastructure (the darker and bolder the stretches of road, the higher the AADT) and the offloading overlay above. The AADT (*Annual Average Daily Traffic*) is the volume of traffic over a year divided by the number of days in the year.

of all requests indexed by all sources and destinations. The leakage tolerance  $0 < \mathcal{L}^{st} \leq 1$  refers to the proportion of data that can be lost during the transfer. Finally, we denote by  $\mathcal{B}^{st}$  the total amount of data (including redundant data) transferred between  $s$  and  $t$ .

We assume that all offloading spots replicate data on multiple outgoing vehicles. For a given request  $\mathcal{R}^{st}$ , each offloading spot  $i$  replicates the data  $\rho^{st}(i, j)$  times when allocated on logical link  $(i, j) \in L^O$  (i.e., the same data is loaded onto  $\rho^{st}(i, j)$  vehicles). Such replication is needed to mitigate the logical link leakage  $\mathcal{L}(i, j)$  and satisfy the leakage tolerance  $\mathcal{L}^{st}$  of the request.

Given a request  $\mathcal{R}^{st}$ , our offloading service is subject to a gain factor  $\gamma^{st}$  and an operational cost  $\alpha_i$ .  $\gamma^{st}$  is the gain factor in terms of cost to send data from  $s$  to  $t$  using our offloading service over the legacy Internet (e.g., a dedicated connection between  $s$  and  $t$ ). Without loss of generality, we assume that  $\gamma^{st}$  is linear with the volume of data transferred (i.e., for one bit we transfer using our offloading service, we save  $\gamma^{st}$  compared to the legacy Internet).  $\alpha_i$  is the operational cost at offloading spot  $i$  to handle the storage of data that is unloaded from incoming vehicles.

In the following, we state our data delivery problem as a linear optimization problem. We define the revenue as the difference between the gain we generate using the offloading service to transfer  $\mathcal{B}^{st}$  data within duration  $\mathcal{T}^{st}$  and the cost of handling the data at the offloading spots. This problem aims at maximizing the revenue, which is constrained by the delay tolerance  $\mathcal{T}^{st}$  and the capacity  $c(i, j)$  of the logical links.

### IV. MAPPING OF THE OFFLOADING OVERLAY

A logical link  $(i, j) \in L^O$  (see Section III-A) is characterized by a travel time  $t(i, j)$ , an aggregated capacity  $c(i, j)$ , and a data leakage  $\mathcal{L}(i, j)$ . In the following, we consider O-D flows between two stations, i.e., flows that go from station  $i \in N^S$  to station  $j \in N^S$ . Let  $r$  be the number of “reasonable” paths between  $i$  and  $j$ , i.e., paths that do not backtrack.<sup>2</sup> The value  $r$  depends on the traffic assignment

<sup>1</sup>The authors are conscious that there are some privacy issues involved in this process, but they are out of the scope of this paper.

<sup>2</sup>A path that does not backtrack is a path that takes a traveler away from origin and closer to destination.

strategy that we consider (e.g., All-or-Nothing, Wardrop's equilibrium, or Dial's stochastic assignments) [7], [8].

#### A. Travel time of logical links

For stretch of road  $(a, b) \in L^R$ , let  $v_{ab}$  be its nominal volume of vehicles (vehicles per unit of time),  $c_{ab}$  be its capacity (vehicles per unit of time), and  $t_{ab}(0)$  be its corresponding travel time at free-flow speed (i.e., when  $v_{ab} = 0$ ). The travel time of  $(a, b)$  is given by the BPR (Bureau of Public Roads) function defined as [9]:

$$t_{ab}(v_{ab}) = t_{ab}(0) \left[ 1 + \alpha \left( \frac{v_{ab}}{c_{ab}} \right)^\beta \right], \quad (1)$$

where  $\alpha$  and  $\beta$  are BPR parameters that depend on the road profile ( $\alpha = 0.15$  minutes and  $\beta = 4.0$  are typically used) [10].

We can deduce from Eq. (1) the travel time of physical path  $p$ , denoted  $t_p$ , which is the sum of all travel times of the stretches of road the path consists of (we do not consider any turning delays at junctions):

$$t_p = \sum_{(a,b) \in p} t_{ab}(v_{ab}). \quad (2)$$

From Equation (2), we deduce an expression of the average travel time  $t(i, j)$  experienced on the  $r$  physical paths between nodes  $i$  and  $j$ , weighted by the road traffic flow  $v_p$  on each path  $p$ :

$$t(i, j) = \frac{\sum_{p \in \mathcal{P}^{ij}} t_p v_p}{r \sum_{p \in \mathcal{P}^{ij}} v_p}, \quad (3)$$

where  $\mathcal{P}^{ij}$  is the set of all simple physical paths between  $i$  and  $j$  (i.e., with no cycles in the path).

#### B. Capacity at logical links

The capacity  $c(i, j)$  of the overlay link  $(i, j) \in L^O$  depends on the sum of the traffic flows  $v_p$  of the simple paths between offloading spots  $i$  and  $j$  (i.e., the number of vehicles per unit of time going from  $i$  to  $j$  on path  $p$ ). It also depends on the market share  $\mathcal{M}$  of the vehicles participating in the offloading service and the storage size  $\mathcal{S}$  on each vehicle:

$$c(i, j) = \mathcal{M} \times \mathcal{S} \sum_{p \in \mathcal{P}^{ij}} v_p. \quad (4)$$

#### C. Leakage of logical links

The leakage  $\mathcal{L}(i, j)$  (comprised between 0 and 1) of logical link  $(i, j) \in L^O$  represents the proportion of data that is lost on logical link  $(i, j)$ . The leakage increases as more vehicles carrying data prematurely exit the road (e.g., the vehicles may exit the highway before reaching the offloading spot or an accident may have occurred).

### V. OFFLOADING SERVICE REVENUE MAXIMIZATION

We denote by  $p$  a logical simple path (i.e., without any cycles) defined as a set of offloading spots connected by logical

links. We denote by  $\mathcal{P}^{st}$  the set of all logical simple paths between  $s$  and  $t$ .

We define the travel time  $t(p)$  experienced on logical path  $p \in \mathcal{P}$  as the sum of the travel times  $t(i, j)$  on the logical links that compose path  $p$  and the waiting time  $t_i$  at each intermediate offloading spot on the path:

$$t(p) = \sum_{(i,j) \in p} (t(i, j) + t_i). \quad (5)$$

We denote by  $f(p)$  the resulting flow (bandwidth) on logical path  $p \in \mathcal{P}^{st}$  for a given request  $\mathcal{R}^{st}$ . For the same request, the transfer of an amount of data  $\mathcal{B}^{st}$  is constrained by the delay tolerance  $\mathcal{T}^{st}$  and does not depend on replication of data. Hence, the delay constraint is the same for both models. The delay to transfer  $\mathcal{B}^{st}$  between  $s$  and  $t$  is expressed as the sum of the duration to transfer this quantity over all logical paths between  $s$  and  $t$  and the average travel time experienced on these paths, weighted by the flow on each path. The delay constraint is expressed as:

$$\frac{\sum_{p \in \mathcal{P}^{st}} f(p) t(p)}{\sum_{p \in \mathcal{P}^{st}} f(p)} + \frac{\mathcal{B}^{st}}{\sum_{p \in \mathcal{P}^{st}} f(p)} \leq \mathcal{T}^{st}. \quad (6)$$

Eq. (6) is not a linear expression, which can lead to some difficulties to solve the linear optimization problem. But it can be rewritten as the following linear expression:

$$\sum_{p \in \mathcal{P}^{st}} f(p) (\mathcal{T}^{st} - t(p)) \geq \mathcal{B}^{st}. \quad (7)$$

In the model we consider, offloading spots store and replicate data. Consider a given request  $\mathcal{R}^{st}$  with leakage tolerance  $\mathcal{L}^{st}$  and  $p \in \mathcal{P}^{st}$  a path between  $s$  and  $t$ . If  $\rho^{st}(i, j) f(p)$  data is transmitted on overlay link  $(i, j) \in p$ ,  $f(p)$  data is received at destination offloading spot  $j$ . Here, the replication factor  $\rho^{st}(i, j)$  is calculated by each offloading spot  $i$  that sends data to remote offloading spot  $j$  on logical link  $(i, j) \in L^O$  as a function of  $\mathcal{L}^{st}$  and  $\mathcal{L}(i, j)$  as follows:

$$\mathcal{L}(i, j)^{\rho^{st}(i, j)} \leq \mathcal{L}^{st} \implies \rho^{st}(i, j) \geq \frac{\log(\mathcal{L}^{st})}{\log(\mathcal{L}(i, j))}. \quad (8)$$

The flow on logical link  $(i, j) \in L^O$  is expressed as the sum of all flows sent by offloading spot  $i$  to offloading spot  $j$  on logical link  $(i, j)$  for data transfer request  $\mathcal{R}^{st}$ . The flow is constrained by the capacity of the logical link  $c(i, j)$ :

$$\sum_{s, t \in \mathcal{R}} \sum_{\substack{p \in \mathcal{P}^{st} \\ p \ni (i, j)}} \rho^{st}(i, j) f(p) \leq c(i, j). \quad (9)$$

The revenue is expressed as the gain of our offloading service compared to an Internet technique (using the gain factor  $\gamma^{st}$ ) minus the operational cost (costs  $\alpha_i$  at each offloading spot  $i$ ). Our objective is to maximize the total revenue, expressed as:

$$\sum_{s, t \in \mathcal{R}} \mathcal{B}^{st} \gamma^{st} - \sum_{i \in \mathcal{N}^O} \alpha_i \sum_{p \ni i} f(p). \quad (10)$$

We note that if Eq. (10) is optimal (i.e., maximized), then Eq. (7) becomes an equality since the quantities  $\mathcal{B}^{st}$  are constrained only in this equation and  $\gamma^{st} > 0$ . Now, if we consider an equality instead of an inequality for the delay constraint Eq. (7), when replacing  $\mathcal{B}^{st}$  with its expression, the total revenue objective can be rewritten as:

$$\sum_{s, t \in \mathcal{R}} \sum_{p \in \mathcal{P}^{st}} f(p) [\gamma^{st} (\mathcal{T}^{st} - t(p)) - \alpha(p)], \quad (11)$$

where  $\alpha(p) = \sum_{i \in p} \alpha_i$ .

In Eq. (11), if  $\gamma^{st} (\mathcal{T}^{st} - t(p)) - \alpha(p)$  is negative, then  $f(p)$  is null (objective function decreases otherwise). Hence, from Eq. (11), we have:

$$t(p) + \frac{\alpha(p)}{\gamma^{st}} \geq \mathcal{T}^{st} \implies f(p) = 0. \quad (12)$$

Since Eq. (12) is a weight on the logical paths (resulting from the weights of the arcs) that depends only on the overlay graph, it allows narrowing the search space of the paths that do not satisfy Eq. (12). Finally, the formulation can be reduced to the total revenue maximization objective that is subjected to the capacity constraint on each logical link  $(i, j) \in L^O$ :

$$\text{Maximize } \sum_{s, t \in \mathcal{R}} \sum_{p \in \mathcal{P}^{st}} f(p) \psi_{st}(p)$$

Subject to

$$\sum_{s, t \in \mathcal{R}} \sum_{\substack{p \in \mathcal{P}^{st} \\ p \ni (i, j)}} \rho^{st}(i, j) f(p) \leq c(i, j) \quad \forall (i, j) \in L^O$$

where  $\psi_{st}(p) = \gamma^{st} (\mathcal{T}^{st} - t(p)) - \alpha(p)$  can be considered as a weight on logical path  $p \in \mathcal{P}^{st}$ .

The amount of data transferred  $\mathcal{B}^{st}$  within the delay tolerance imposed by request  $\mathcal{R}^{st}$  is deduced from Eq. (7):

$$\mathcal{B}^{st} = \sum_{p \in \mathcal{P}^{st}} f(p) (\mathcal{T}^{st} - t(p)). \quad (13)$$

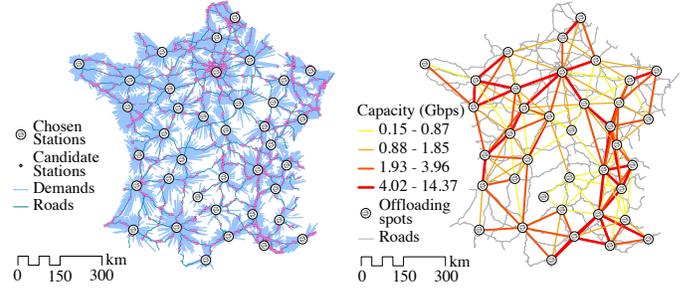
We can finally obtain from Eq. (13) the average bandwidth of request  $\mathcal{R}^{st}$  by dividing the amount of data  $\mathcal{B}^{st}$  by the duration of the transfer  $\mathcal{T}^{st}$ .

## VI. EVALUATION

We evaluate our offloading service model with real traffic datasets measured in France's road infrastructure.

### A. Dataset and planning of the network of stations

We use road traffic data collected in 2011 on the major stretches of road in France [11]. The dataset shows AADT (*Annual Average Daily Traffic*), i.e., the total volume of traffic measured on both directions of the stretches of road for a year divided by the number of days in the year. The graph is composed of 3,310 edges covering over 20,000 km of roads.



(a) Allocation of stations over the French road infrastructure. (b) Offloading overlay with the average capacity of overlay links.

Fig. 2: Facility-allocation result and offloading overlay. The big dots are the chosen stations.

We consider a network of stations equivalent to the one Tesla is currently rolling out in North America [5]. Since there is no such a network in France for the time being, we plan a simple yet realistic network of stations over the road infrastructure of France. We consider a facility-allocation problem that minimizes the number of facilities to allocate, a problem we adapted from the maximal covering location problem [12]. The problem takes demand points and candidate locations as inputs: (i) the demand points are the 9,555 cities of France with a population greater than 1,000, and (ii) the candidate locations are the 1,024 gas stations of Total, a major oil company in France.

The facility-allocation algorithm chooses the stations such that maximum demand points are allocated to the stations within a range of 150 km, while minimizing the number of chosen stations. The resulting allocation gives 38 stations scattered on the French road infrastructure as shown in Fig. 2a.

### B. Mapping of the offloading overlay

To compute the overlay depicted in Fig. 2b, we consider the All-or-Nothing traffic assignment strategy that assigns all the traffic between a source and a destination to the shortest path, here defined as the path with the shortest travel time [7], [8]. The traffic flow on this path is set to the minimum AADT of the stretches of road composing it. Then, we use the results of the most recent French travel household survey made in 2008 (ENTD), which distinguishes two kinds of travel [13]: local travel is less than 80 km from home, while long-distance travel is greater than 80 km from home. The latter one accounts for 2.8% of the total amount of recorded travel. We consider a market share  $\mathcal{M}$  of 20% (which is the case for Renault in France) and a per-node storage unit  $\mathcal{S}$  of one TB (assumed to be the same for each electric vehicle).

Since the AADT is measured on both directions of a stretch of road, we divide the measured values by two. To compute the bandwidth of the logical link, we multiply the resulting flow by  $\mathcal{S}$ . We use the BPR (Eq. (1)) to calculate the travel time on each stretch of road with:  $\alpha = 0.15$  minutes,  $\beta = 4.0$ , the practical capacity given by the dataset, and the AADT for the assigned volume. We use then Eq. (3) to deduce the travel time of the logical links using the shortest physical road paths.

Fig. 3 shows the benefits of using the offloading overlay

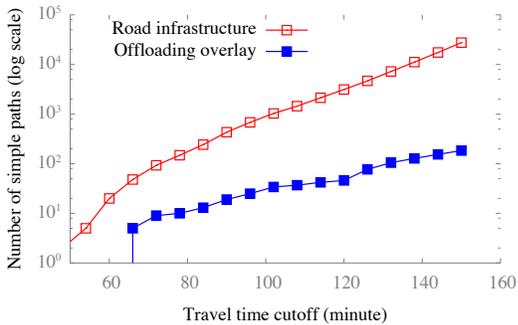


Fig. 3: Total number of simple paths in the road infrastructure and logical simple paths in the offloading overlay as a function of the travel time cutoff ( $y$ -axis is in a logarithmic scale).

compared to the road infrastructure. We note that the number of simple paths grows exponentially with the travel time cutoff and is much larger on the road infrastructure. Also, the difference in the number of paths grows exponentially.

### C. Data transfer assignment problem

We interface our offloading overlay with IBM ILOG CPLEX 12.5.1 using the linear optimization models we presented in Section V. We consider a scenario with the following three different requests (distances are Euclidean):

- Request R1: from Paris to Lyon (384 km).
- Request R2: from Paris to Bordeaux (492 km).
- Request R3: from Paris to Marseille (646 km).

It is important to note that requests R1 and R3 will compete for the flows since they share some common subpaths. We use a breadth-first search algorithm to generate all the simple paths for each request. The cutoffs of the simple paths are given by Eq. (12). However, the generation of the simple overlay paths between a source and a destination is exponential, as seen in Fig. 3. To solve this issue, we reduce our search space by applying a default cutoff of 12 hours on the travel time of the simple logical paths we generate for our experiments.

We express the gain factor  $\gamma^{st}$  as an exponential function of the Euclidean distance (in km) between  $s$  and  $t$ :  $\gamma^{st} = [\text{dist}(s, t)]^\beta$ , where  $\beta \in \mathbb{R}$ . In our analyses, we investigate the impact of the delay tolerance  $\mathcal{T}^{st}$ . By default, for all requests, we set leakage tolerance  $\mathcal{L}^{st}$  to  $10^{-2}$  and link leakage  $\mathcal{L}(i, j)$  to 0.05 for each logical link  $(i, j) \in L^O$ . The operational cost  $\alpha_i$  is weighted by the demands allocated to offloading spot  $i$ , resulting from the facility-allocation problem. Since we want a fair flow allocation in our offloading service, we choose  $\beta$  such that the standard deviation of the bandwidth of all flows is minimized. Finally, for each offloading spot  $i$ , we consider a waiting time  $t_i = 20$  minutes, which corresponds to the duration of a charge that provides a 300-km range to the vehicles [5].

Fig. 4 shows the average bandwidth of each request as a function of the delay tolerance  $\mathcal{T}^{st}$ . We notice that the bandwidth stabilizes as the duration of the transfer increases. Also, requests with longer distances are favored (request R2 and R3) over requests with shorter distances (request R1) since the total revenue is increased.

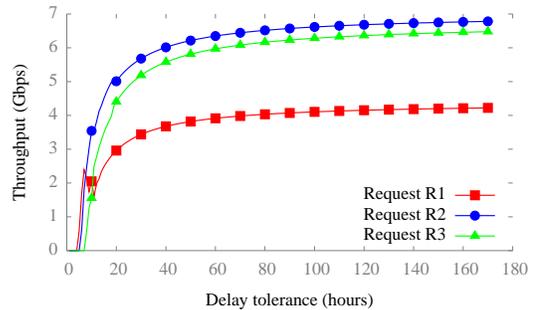


Fig. 4: Bandwidth as a function of delay tolerance  $\mathcal{T}^{st}$

## VII. CONCLUSION

We proposed a bulk-data transfer service to offload delay-tolerant Internet traffic onto the road infrastructure network connecting geographically distant offloading spots. We evaluated our system using real traffic numbers in France. Our offloading service allows data transfers in the petabyte range per week while achieving cumulative bandwidth exceeding 10 Gbps with a market share of 20% and only one terabyte of storage per vehicle. These results confirm the offloading potential of our service which can help operators handle big data, especially during busy periods.

As future work, we plan to replicate the data partially at some offloading spots or at the source offloading spot. We also intend to address security and privacy concerns, as well as the adaptation of our models so as to capture diurnal and seasonal variations of the road traffic.

## ACKNOWLEDGMENT

Benjamin Baron, Promthe Spathis, and Marcelo Dias de Amorim carried out part of the work at LINCOS (<http://www.lincos.fr>).

## REFERENCES

- [1] Cisco Visual Networking Index, “The zettabyte era—trends and analysis,” *Cisco white paper*, 2013.
- [2] A. Keränen and J. Ott, “DTN over aerial carriers,” in *ACM Mobicom workshop on challenged networks (CHANTS)*, Beijing, China, Sep. 2009.
- [3] B. Cho and I. Indranil, “Budget-constrained bulk data transfer via internet and shipping networks,” in *ACM International Conference on Autonomic Computing*, Karlsruhe, Germany, Jun. 2011.
- [4] D. Waitzman, “A standard for the transmission of ip datagrams on avian carriers,” Internet Requests for Comments, RFC Editor, RFC 1149, April 1990. [Online]. Available: <http://tools.ietf.org/html/rfc1149>
- [5] Tesla Motors, Inc, “Supercharger,” <http://www.teslamotors.com/supercharger>.
- [6] C.-Y. Hong, S. Kandula, R. Mahajan, M. Zhang, V. Gill, M. Nanduri, and R. Wattenhofer, “Achieving high utilization with software-driven wan,” in *ACM Sigcomm*, Hong Kong, China, Aug. 2013.
- [7] J. de Dios Ortúzar and L. Willumsen, *Modelling transport*. Wiley Chichester, 2011, vol. 7.
- [8] Y. Sheffi, *Urban transportation networks: equilibrium analysis with mathematical programming methods*. Prentice Hall, 1985.
- [9] United States. Bureau of Public Roads, *Traffic assignment manual for application with a large, high speed computer*, 1964.
- [10] National Research Council (U.S.). Transportation Research Board, *Highway Capacity Manual*, 2000.
- [11] DGITM/DIT – SETRA – IGN, “Recensement de la circulation sur le réseau routier national en 2011,” <http://tinyurl.com/otfbewv>.
- [12] R. Church and C. R. VELLE, “The maximal covering location problem,” *Papers in regional science*, 1974.
- [13] MEDDE, “Mobilité – déplacement,” <http://tinyurl.com/otfbewv>.