An efficiency measure for road transportation networks with application to two case studies

Håkan Terelius
Karl Henrik Johansson

Abstract—Enabling efficient transportation is a major challenge for large cities, as the transportation need is increasing, while the environmental impact has to be minimized. In this paper, we define an efficiency measure that shows how much of the current transportation mileage that is really necessary to meet all the transportation assignments. We show that the efficiency measure can be computed efficiently as a minimum cost flow, and we apply it on two case studies. The first case demonstrate the efficiency measure on a freight transportation system, and the second case computes the measure for a large real-world data set from the New York City taxis.

I. Introduction

Transportation of both goods and people is essential for the function of our society. The transportation system is also facing great challenges, as the demand is steadily increasing, while the cost and environmental impacts needs to minimized. In 2012, 26% of all greenhouse gas emissions from the European Union could be derived from the transportation sector, and 18% for road transportation alone [1].

The transportation flow and congestion problem in the transportation system has been studied for over 80 years [2]. The distribution problem between a set of origins and a set of destinations was formulated in [3], and spawned a wide research in optimal flow allocation [4, 5, 6].

One challenge for freight carriers is the need to move empty vehicles to avoid an accumulation of empty vehicles in a region, known as dead mileage. Studies have estimated that up to 40% of both the mileage and cost in different transportation systems are due to empty vehicles [7]. Optimizing fleet management systems is therefore actively pursued [8], but many of the combinatorial optimization problems are NP-complete, and therefore intractable to solve, such as the traveling salesman problem [9].

Current research in intelligent transportation systems aim to provide more information and better decisions in the transportation network. A long-term goal is to provide a completely autonomous transportation system, thereby increasing the safety and efficiency [10]. One research direction is toward vehicle platooning [11], but it is not only the vehicles that are changing, we are also seeing developments in efficient and intelligent infrastructure, such as ride-sharing lanes [12] and reversible lanes [13].

Over the last decades, technology development has enabled a widespread adoption of GPS receivers for determining the position of vehicles, and smart-phones has made it easy to share and collect this position data. This means that today we have access to huge datasets of trajectories from past transportation assignments, and this has opened up new opportunities for understanding transportation patterns [14, 15]. Authorities in many cities have collected and provided extensive dataset from the taxi system, which has stimulated research in this field [16]. For example, [17, 18, 19, 20] studied strategies for cruising taxis and dispatch systems, while [21] estimated the efficiency of a taxi system using a model of perfect prior knowledge of the demand.

The main contribution of this paper is to introduce a new efficiency measure for transportation flows. It analyzes large datasets of trip trajectories, and determines how much the dead mileage could be reduced in an ideal setting where the actors collaborate. We show that it is computationally efficient, and apply it on two case studies. Future research directions includes on-line optimization of the efficiency measure to improve fleet management systems and taxi dispatch systems.

The remainder of this paper is organized as follows: In section II, we introduce a new efficiency measurement for transportation systems, and present some of it properties. In section III we demonstrate the efficiency measure on a freight transportation system, and in section IV we compute the efficiency measure on a huge real dataset from the New York City taxi system. some final conclusions are given in section V.

II. Transportation Efficiency

In this section, we define the transportation efficiency measure, and show some of its properties. The aim is to measure the efficiency of a transportation system consisting of homogeneous actors, compared to the idealized case where all actors are collaborating to satisfy the demands.

A. Network Model

Given a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of nodes, and $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ is the set of directed edges. For each edge $(u, v) \in \mathcal{E}$, there are two associated transportation flows $f_1 : \mathcal{E} \to \mathbb{R}_+$, and $f_0 : \mathcal{E} \to \mathbb{R}_+$. The first flow, f_1, represents the desired *transportation assignments*, while the second flow, f_0, represents the vacant trips taken in order to move the vehicles to their next transportation assignments, which we refer to as *vacant flow*. Each edge $(u, v) \in \mathcal{E}$ also has an associated weight $w : \mathcal{E} \to \mathbb{R}_+$, which is the cost of transporting one unit of flow across the edge.

B. Transportation Efficiency Measure

In this paper, we consider a measure for the efficiency of the transportation system. First, let us define the network flow cost C as the total cost for all trips (both the transportation assignments and the vacant flow),

$$C \doteq \sum_{(u,v) \in \mathcal{E}} w(u,v) \left(f_0(u,v) + f_1(u,v) \right).$$

Now, let us consider an optimal network flow by reducing the vacant flow f_0 while preserving the availability of vacant vehicles as follows

$$f_0^\star \doteq \arg\min_f \sum_{(u,v) \in \mathcal{E}} w(u,v) \left(f(u,v) + f_1(u,v) \right)$$

subject to

$$0 \le f(u,v) \le f_0(u,v), \quad \forall (u,v) \in \mathcal{E},$$

and

$$\sum_{\substack{v \in \mathcal{N} \\ (u,v) \in \mathcal{E}}} f(u,v) - \sum_{\substack{v \in \mathcal{N} \\ (v,u) \in \mathcal{E}}} f(v,u)$$
$$= \sum_{\substack{v \in \mathcal{N} \\ (u,v) \in \mathcal{E}}} f_0(u,v) - \sum_{\substack{v \in \mathcal{N} \\ (v,u) \in \mathcal{E}}} f_0(v,u), \quad \forall u \in \mathcal{N}.$$
$$(1)$$

The first constraint implies that the optimal flow $f_0^\star$ is a subset to the vacant flow f_0, i.e., it does not increase the flow over any edge. The second constraint implies that the excess flow at every node is preserved, i.e., that the optimal flow leaves the same number of vacant vehicles at every node for their next assignments.

We can now define the optimal network flow cost as

$$C_{\text{opt}} \doteq \sum_{(u,v) \in \mathcal{E}} w(u,v) \left(f_0^\star(u,v) + f_1(u,v) \right),$$

and we are now ready for the main definition in this paper.

Definition 1. *The **transportation efficiency measure** is defined as*

$$\eta \doteq \frac{C_{opt}}{C}.$$

Remark 1. *We have $\eta \in [0,1]$, and $\eta = 1$ if the transportation system is optimal. A value $\eta < 1$ shows*

how inefficient the system is, as it measures the percentage of the trips that are actually necessary to fulfill all the transportation assignments with a given vehicle fleet.

Remark 2. *Since the optimal flow $f_0^\star$ only reduces the initial vacant flow f_0, we neglect any changes in congestion that would appear from an increased traffic flow. Thus, computing this efficiency measure η can be done directly from historical GPS trajectories, as seen in section IV.*

Remark 3. *We do not assume that the actual transportation assignments are known, but only the historical traces of the vehicle trajectories. Hence, it is natural to assume that all transportation assignments f_1 has to be fulfilled, and that they also represent the complete transportation demand.*

Remark 4. *Time constraints on the transportation assignments are not directly captured by this flow model. However, a possible solution is to only consider the trips that occur during a limited time period, and then recompute the efficiency measure for each time period.*

Remark 5. *Even though the entire transportation system is not homogeneous, we can consider a subsystem with homogeneous actors, e.g., a set of long haul trucking companies or a set of taxi drivers, which could benefit from collaboration.*

C. Example

As a simple example, consider a scenario with three companies (Blue, Red and Green) moving cargo between the three cities (A, B and C). Company Blue moves cargo from city B to city A, and returns empty to pickup the next cargo. Company Red similarly moves cargo from city C to city B, and returns empty, while company Green moves cargo from city A to city C, and returns empty, as illustrated in fig. 1.

With a unit cost $w(u, v) = 1$ for all edges $(u, v) \in \mathcal{E}$, the total cost is $C = 6$. Notice that in this example, there is a cycle of empty trucks going around from A to B to C, and back to A, and that the transportation assignments could be served by a single truck going around from A to C to B and back to A, as shown in fig. 2.

The cost for this optimized network is $C_{\text{opt}} = 3$, thus the efficiency of the transportation system is only $\eta = \frac{C_{\text{opt}}}{C} = \frac{3}{6} = 50\%$.

D. Computational Complexity

In this section, we show that the efficiency measure can be computed efficiently. In many practical scenarios, the road network consist of thousands of nodes, and there can be billions of collected transportation trips, as is illustrated in section IV. Therefore it is essential that η can be efficiently computed.

The major computational step is to solve the optimal network flow problem in eq. (1), which we will show is equivalent to the *minimum cost flow* problem. Recall the
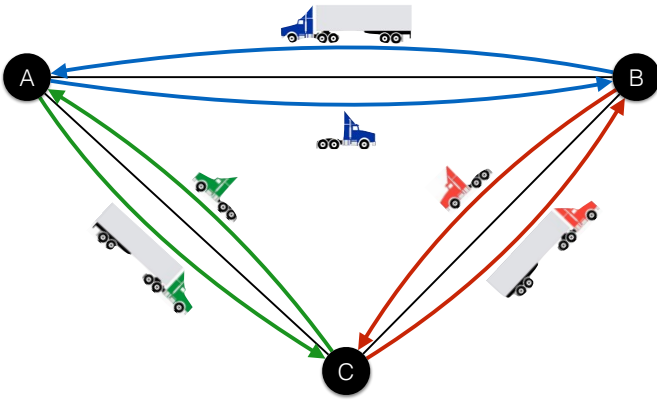
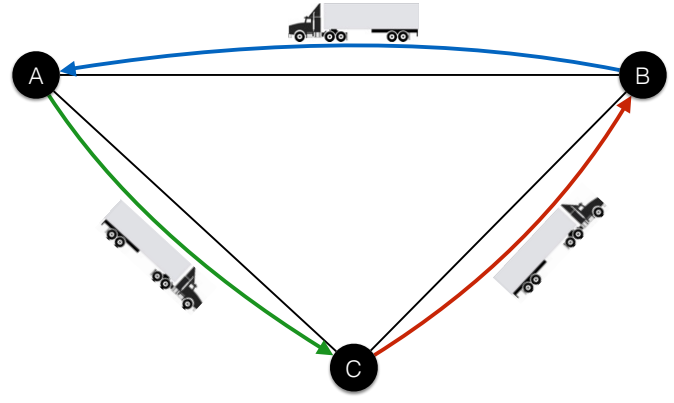Figure 1. Example of a three company transportation network.

Figure 2. Optimized transportation network served by a single truck.

minimum cost flow problem formulation [22], which can be stated as

$$\min_{f} \sum_{(u,v) \in \mathcal{E}} w(u,v) f(u,v)$$

subject to

$$0 \leq f(u,v) \leq c(u,v), \quad \forall (u,v) \in \mathcal{E},$$

and

$$\sum_{\substack{v \in \mathcal{N} \\ (u,v) \in \mathcal{E}}} f(u,v) - \sum_{\substack{v \in \mathcal{N} \\ (v,u) \in \mathcal{E}}} f(v,u) = b_u, \quad \forall u \in \mathcal{N},$$

(2)

where $c(u,v)$ is the edge capacity, and b_u is the node supply/demand.

It is furthermore assumed by the *feasibility assumption* that $\sum_{u \in \mathcal{N}} b_u = 0$, and that there exist a feasible solution to eq. (2).

Clearly, with the capacity constraint given by $c(u,v) = f_0(u,v)$ for all edges $(u,v) \in \mathcal{E}$, and the supply/demand constraint given by $b_u = \sum_{\substack{v \in \mathcal{N} \\ (u,v) \in \mathcal{E}}} f_0(u,v) - \sum_{\substack{v \in \mathcal{N} \\ (v,u) \in \mathcal{E}}} f_0(v,u)$ for all nodes $u \in \mathcal{N}$, our problem in eq. (1) is of the same form as eq. (2). Further, by rearranging the sums, we see that $\sum_{u \in \mathcal{N}} b_u = \sum_{u \in \mathcal{N}} \sum_{\substack{v \in \mathcal{N} \\ (u,v) \in \mathcal{E}}} f_0(u,v) - \sum_{u \in \mathcal{N}} \sum_{\substack{v \in \mathcal{N} \\ (v,u) \in \mathcal{E}}} f_0(v,u) = \sum_{u \in \mathcal{N}} \sum_{\substack{v \in \mathcal{N} \\ (u,v) \in \mathcal{E}}} f_0(u,v) - \sum_{u \in \mathcal{N}} \sum_{\substack{v \in \mathcal{N} \\ (u,v) \in \mathcal{E}}} f_0(u,v) = 0$. Finally, it is easy to see that $f = f_0$ is a feasible solution, thus we can conclude that we have a feasible minimum cost flow problem.

A complexity survey of minimum cost flow algorithms is presented in [23], where they show that the *generalized cost-scaling algorithm with dynamic trees* is one of the fastest algorithms, with time complexity $O\left(nm \log(n^2/m) \min\{\log(nW), m \log n\}\right)$. Here, $n = |\mathcal{N}|$ is the number of nodes, $m = |\mathcal{E}|$ is the number of edges, and $W = \max_{(u,v) \in \mathcal{E}} w(u,v)$ is the largest edge weight.

We can summarize this in the following proposition.

Proposition 1. *The transportation efficiency measure η can be computed as a minimum cost flow in polynomial*

time, e.g., $O\left(nm \log(n^2/m) \min\{\log(nW), m \log n\}\right)$.

Remark 6. *There are many other algorithms for solving the minimum-cost flow problem. We used the* successive shortest path algorithm *[6] successfully for the following case studies, with a theoretical worst case performance $O\left(D(m + n \log n)\right)$, where D is the maximum flow value with integer capacities.*

Remark 7. *As shown in [24], solving the minimum cost flow problem is equivalent to removing all negative weight cycles from a feasible solution. This can be interpreted in our application as removing all cycles traveled by the empty vehicles. Compare with figs. 1 and 2.*

III. Freight Transportation Case Study

In this section, we demonstrate the transportation efficiency measure η on a simulated road transportation system. The transportation system consists of competitive transportation companies that receives transportation assignments in a discrete time model. Each company optimizes the routes of its own vehicles, without any collaboration with the other companies.

The road network is based on the OpenStreetMap data [25] for the German highways, shown in fig. 3. Germany's 14 largest cities are selected as the nodes, and the weight w on each edge is set as the travel distance between the two cities.

We consider 10 transportation companies, each of which has $T = 20$ trucks in their vehicle fleet. In a discrete time model, at each time step, each company receives a set of T randomized transportation assignments, where the probability distribution for the pickup and drop-off locations is proportional to the population size of the cities. Each time step is divided into two phases, first the vehicles are moved from their current location to the pickup location of the transportation assignment, and in the second phase, the transportation assignments are carried out. Each company optimizes its allocation of vehicles to transportation assignments in order to minimize
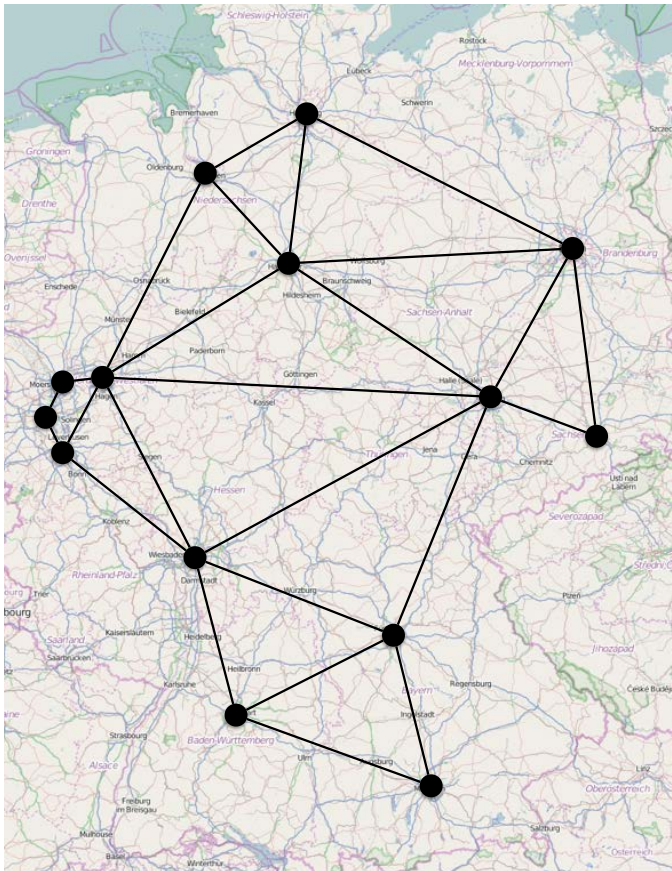
Figure 3. Road network between Germany's 14 largest cities. Map courtesy of OpenStreetMap.

the total dead mileage for its own vehicle fleet, using a weighed bipartite matching. These trips constitutes the vacant flow f_0, and transportation assignment trips in the second phase constitutes f_1. At the end of each time step, the vehicles are left at the drop-off locations, waiting for the next time step.

At each time step, we thus defined the two network flows f_0 and f_1, and can compute the optimal flow $f_0^\star$ and transportation efficiency measure η. This procedure is repeated (for 1000 time steps) until the efficiency measure converges, and the results are shown in table I. The transportation efficiency measure is $\eta = 87\%$, which implies that in this case study, where each company optimized its own fleet, 13% of the total mileage could be eliminated if the companies would start to collaborate with each other.

Table I
FREIGHT TRANSPORTATION RESULTS WITH 20 TRUCKS PER COMPANY.

Avg. assignment length	454.7 km
Avg. original dead mileage	125.1 km
Avg. optimized dead mileage	51.9 km
Transportation efficiency η	87%

In fig. 4, we change the number of trucks per company T, while keeping the other parameters as before. As can

be seen, the efficiency of the system increases with larger companies in this scenario, since a larger company has more options for optimizing its own routes. However, notice that the total mileage is also increasing, so even when the system efficiency is increasing, the absolute savings for collaborating does also increase.
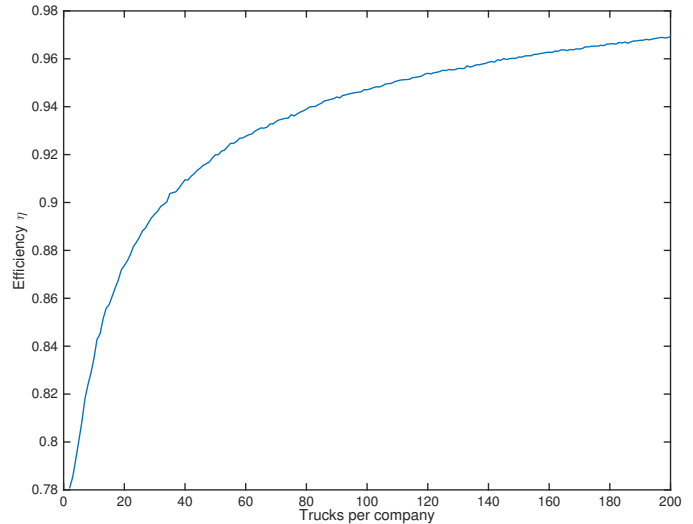

Figure 4. Freight transportation efficiency depending on the number of trucks per company.

IV. NEW YORK TAXI CASE STUDY

We now turn to a second illustration of the efficiency measure η, computed on a real dataset from the New York City taxis. We will show how the efficiency of the taxi system varies depending on the time of the day.

A. Data format

In New York City, the taxi system consists of 13000 yellow medallion taxis, which completed more than 174 million trips during 2013. Records from all these trips has been made publicly available by the NYC Taxi & Limousine Commission. Each trip is specified with the following fields:

- Car ID
- Driver ID
- Pickup time
- Drop-off time
- Passenger count
- Trip distance
- Pickup location (GPS position)
- Drop-off location (GPS position)

A majority of the trips are centered on Manhattan Island, as shown by fig. 5.

B. Data Processing

To compute the efficiency measure η we first need to construct the flow network from the taxi data. This procedure is described in the next four subsections.
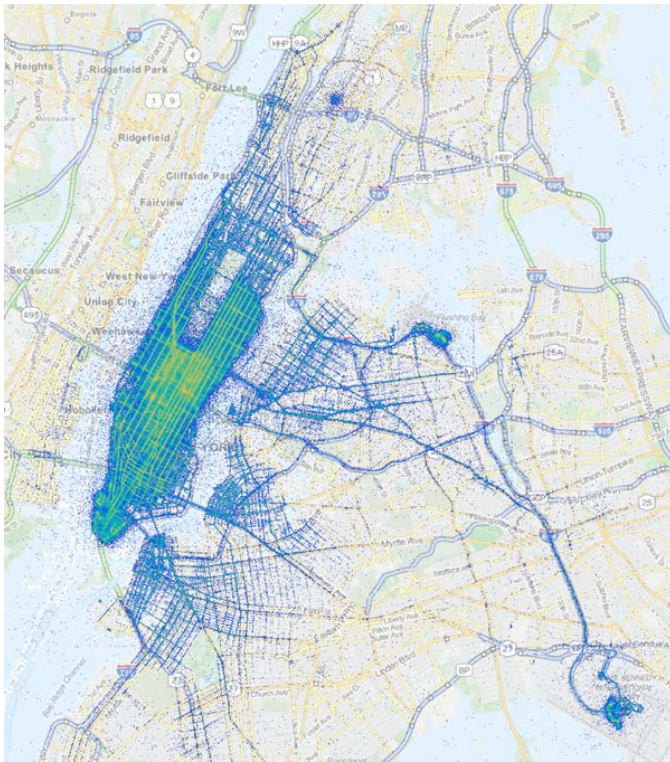
Figure 5. Heat map of the New York City taxis' pickup locations. Map courtesy of OpenStreetMap.

Figure 6. The average number of trips taken as a function of the time of day.

1) *Determining the vacant flow :* The dataset contains the transportation assignments, where the taxi is driving with passengers, but does not include the cruising trips where the driver is looking for passengers. However, since the transportation assignments are specified with both a car identity, a driver identity, as well as the time of the trip, we can determine the next trip for each car and driver. If a car-driver pair drove another trip within one hour, then we add the empty trip from the previous trip's drop-off location to the next trip's pickup location.

For 91.3% of all transportation assignments, we could find a following vacant flow trip. The average transportation assignments and the vacant flow variation over a day is shown in fig. 6.

Remark 8. *There is a dip in the number of available taxis in the middle of the afternoon rush, because the drivers traditionally change shifts at this time [26].*

2) *OpenStreetMap network :* The trip data only contains the pickup and drop-off locations, therefore we construct a network based on the road data from OpenStreetMap [25]. The map region around New York contains 1460536 nodes and 2967562 edges, and the pickup and drop-off locations for each trip are mapped to the closest nodes in the OpenStreetMap data.

The next step is to find the path through the road network for every trip. To this end, we compute the shortest path for each trip, taking into account the road
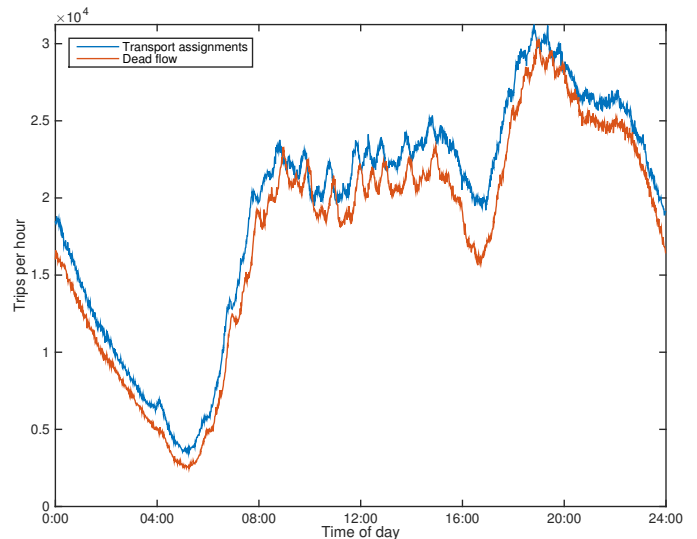
type, speed limits and one-way directions. The result is a prediction of how the taxis are moving, similar to the suggestions given by GPS navigators.

3) *Grid regions :* The OpenStreetMap data contains a very detailed road network, including many local phenomena, e.g., complex intersections, parallel lanes, antiparallel one-way streets, etc. But we are only interested in the general trajectories of the taxis, which would be hidden by excessive details of the map.

To address this problem, we divide the map into a square grid, where we vary the cell sizes from $100\,\mathrm{m} \times 100\,\mathrm{m}$ up to $5000\,\mathrm{m} \times 5000\,\mathrm{m}$. Each grid cell then becomes a node in our final flow network, and a taxi trip is represented as a sequence of adjacent grid cells. Because the grid cells have the same size, we use a unit weight $w(u, v) = 1$ for all $(u, v) \in \mathcal{E}$, and the flow is equal to the number of taxis passing between two grid cells.

Remark 9. *This has the additional benefit of reducing the number of nodes in the flow network, which makes the computations faster.*

4) *Computing the efficiency measure :* We have now constructed the flow network from the pickup and drop-off locations of the taxis, and from this description we are able to compute the efficiency measure η.

Given a year's worth of data, we introduce time slots, where the length is varied from 1 minute up to 60 minutes. For each time slot, we compute the efficiency measure η, and in fig. 7 the taxi transportation efficiency is shown as a function of the time of the day. In table II, the efficiency is shown for different grid sizes and time slots.

C. Results

The average efficiency measure over January month 2013 is shown in table II. The efficiency of the NYC taxi

Table II
NEW YORK CITY TAXI EFFICIENCY RESULTS

| | | | Time slot | | | |
			1 min	5 min	60 min	1 month
Grid size	100 m × 100 m	Avg. flow / time slot	25409	127047	1524573	1132757899
		Dead flow	28.1%	28.1%	28.1%	27.3%
		Efficiency measure η	90.7%	86.6%	84.7%	80.3%
	500 m × 500 m	Avg. flow / time slot	4724	23620	283442	210597439
		Dead flow	27.6%	27.5%	27.5%	26.8%
		Efficiency measure η	88.2%	85.5%	84.4%	80.6%
	1000 m × 1000 m	Avg. flow / time slot	2222	11111	133340	99072081
		Dead flow	27.6%	27.6%	27.6%	26.9%
		Efficiency measure η	87.4%	85.1%	84.2%	80.6%
	5000 m × 5000 m	Avg. flow / time slot	346	1730	20763	15427468
		Dead flow	28.4%	28.4%	28.4%	27.8%
		Efficiency measure η	86.1%	84.5%	83.8%	81.2%

system varies from 90.7% using 1 minute time slots and a grid of 100 m × 100 m, down to 83.8% for 60 minutes time slots and 5000 m × 5000 m grid size.

Notice that the efficiency drops when the time slots and grid size increases. This supports our intuition, because by increasing the time slots, we consider more vehicles at each step, and are therefore more likely to find vehicles that can be eliminated. Similarly for the grid size, a larger grid means that more roads will be part of the same flow edge, e.g., when two cars are traveling on parallel one-way streets, in opposite directions.

Remark 10. *The time slot and grid size can be interpreted as how close two taxis need to be in time and space in order to be considered as redundant.*

In fig. 7, the efficiency is shown as a function of the time of day. Again, we see that a larger time slot yields a consistently lower efficiency. Notice that the efficiency has a peak during the afternoon rush at 5PM. Comparing with fig. 8, we see that this peak corresponds to a very low vacant flow, i.e., the high demand for taxis makes it easy to pick up new passengers. In contrast, the morning peak at 4AM in efficiency corresponds to a high percentage of vacant flow. Comparing also with fig. 6, this is explained by that much fewer taxis are available in the morning due to the low demand for taxis.

V. CONCLUSIONS

In this paper, we developed a new efficiency measure η for transportation systems, where the trips are divided into actual transportation assignments and vacant trips. The efficiency measure is especially useful when evaluating a transportation system based on collected GPS trajectories from the vehicles, and we show that the efficiency measure can easily be computed, even for huge datasets.

We used the efficiency measure to evaluate a realistic road transportation simulation, with multiple competitive transportation companies. We showed that even if the companies are optimizing their own transportation routes, there is still a potential of reducing the total mileage by 13% if they would start to cooperate.
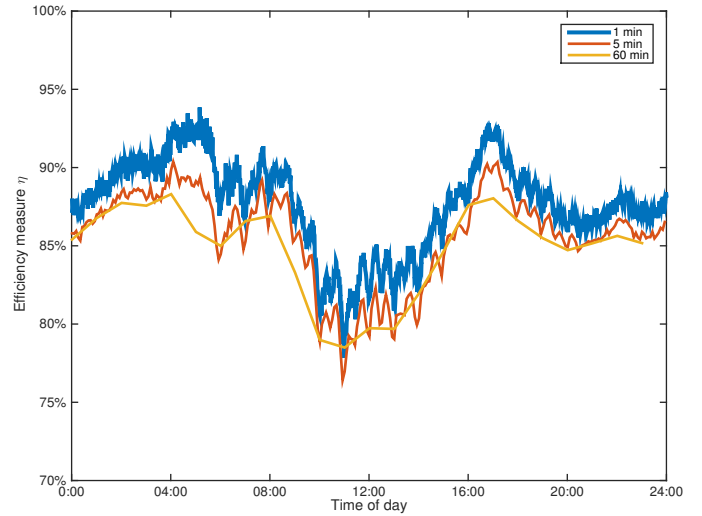

Figure 7. The average efficiency measure by the time of day, using a grid size of 1000 m × 1000 m.
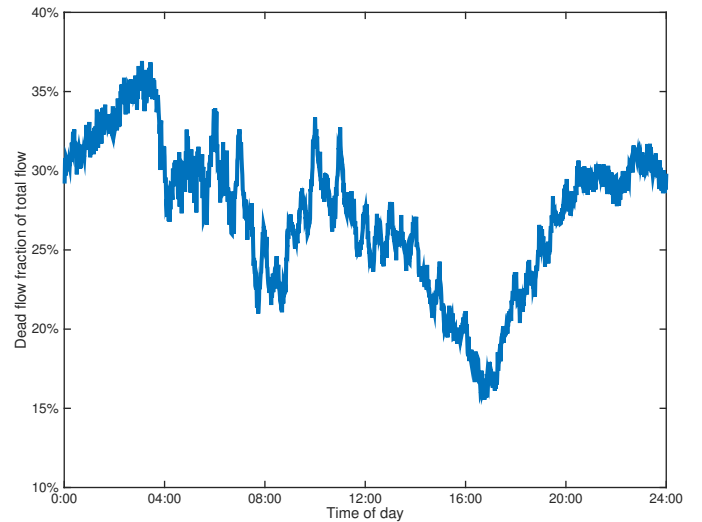

Figure 8. The fraction $f_0/(f_0 + f_1)$ of vacant flow compared to the total flow.

Finally, we also demonstrate the efficiency measure on

a real data set from New York's taxis, where we showed that the total mileage could be reduced by between 9% and 20%, depending on the model.

This work shows that there is a huge potential for improving the efficiency of our current transportation system by improved planning and coordination among different actors. The efficiency measure can also be used in on-line settings, to evaluate and optimize different transportation strategies. In the freight transportation case, the optimization can be performed either in a fleet owner's logistic system, or provided by a third-party for collaboration between multiple fleet owners.

Future work includes studying different policy changes for improving the transportation efficiency.

References

[1] European Union, *EU Transport in Figures*. European Union, 2014.

[2] B. Greenshields, J. Bibbins, W. Channing, and H. Miller, "A study of traffic capacity," *Highway research board proceedings*, vol. 14, pp. 448–477, 1935.

[3] F. L. Hitchcock, "*The distribution of a product from several sources to numerous localities*," *Journal of Mathematical Physics*, no. 20, pp. 224–230, 1941.

[4] T. C. Koopmans, "Optimum Utilization of the Transportation System," *Econometrica*, vol. 17, pp. 136–146, 1949.

[5] L. R. Ford and D. R. Fulkerson, "A simple algorithm for finding maximal network flows and an application to the Hitchcock problem," *Canadian Journal of Mathematics*, pp. 210–218, 1957.

[6] J. Edmonds and R. M. Karp, "Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems," *Journal of the ACM*, vol. 19, no. 2, pp. 248–264, Apr. 1972.

[7] P. J. Dejax and T. G. Crainic, "A Review of Empty Flows and Fleet Management Models in Freight Transportation," *Transportation Science*, vol. 21, no. 4, pp. 227–248, Nov. 1987.

[8] T. G. Crainic and G. Laporte, Eds., *Fleet Management and Logistics*. Springer, 1998.

[9] M. R. Garey and D. S. Johnson, *Computers and Intractability*, ser. A Guide to the Theory of NP-Completeness. W H Freeman & Company, Jan. 1979.

[10] O. Gusikhin, D. Filev, and N. Rychtyckyj, "Intelligent Vehicle Systems: Applications and New Trends," in *Informatics in Control Automation and Robotics*. Berlin, Heidelberg: Springer, 2008, pp. 3–14.

[11] J. Mårtensson, A. Alam, S. Behere, M. A. A. Khan, J. Kjellberg, K.-Y. Liang, H. Pettersson, and D. Sundman, "The Development of a Cooperative Heavy-Duty Vehicle for the GCDC 2011: Team Scoop," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1033–1049, 2012.

[12] J. Kwon and P. Varaiya, "Effectiveness of California's High Occupancy Vehicle (HOV) system," *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 1, pp. 98–115, Feb. 2008.

[13] Z. Bede, G. Szabó, and T. Péter, "Optimization of road traffic with the applied of reversible direction lanes," *Periodica Polytechnica Transportation Engineering*, vol. 38, no. 1, pp. 3–8, 2010.

[14] G. Gidófalvi and T. B. Pedersen, "Mining Long, Sharable Patterns in Trajectories of Moving Objects," *GeoInformatica*, vol. 13, no. 1, pp. 27–55, Dec. 2007.

[15] Y. Matsubara, L. Li, E. Papalexakis, D. Lo, Y. Sakurai, and C. Faloutsos, "F-Trail: Finding Patterns in Taxi Trajectories," in *17th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Apr. 2013, pp. 86–98.

[16] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban Computing: Concepts, Methodologies, and Applications," *Transactions on Intelligent Systems and Technology*, vol. 5, no. 3, Sep. 2014.

[17] C. Wang, W. K. Ng, and H. Chen, "From data to knowledge to action: A taxi business intelligence system," in *15th International Conference on Information Fusion*, 2012, pp. 1623–1628.

[18] L. Moreira-Matias, R. Fernandes, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "An online recommendation system for the taxi stand choice problem (Poster)," in *Vehicular Networking Conference*. IEEE, 2012, pp. 173–180.

[19] J. Lee, I. Shin, and G.-L. Park, "Analysis of the Passenger Pick-Up Pattern for Taxi Location Recommendation," in *4th International Conference on Networked Computing and Advanced Information Management*. IEEE, Sep. 2008, pp. 199–204.

[20] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-Finder: A Recommender System for Finding Passengers and Vacant Taxis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 10, pp. 2390–2403, Oct. 2013.

[21] X. Zhan, X. Qian, and S. V. Ukkusuri, "Measuring the Efficiency of Urban Taxi Service System," in *UrbComb '14*, Aug. 2014.

[22] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows: theory, algorithms, and applications*. Prentice Hall, 1993.

[23] P. Kovács, "Minimum-cost flow algorithms: an experimental evaluation," *Optimization Methods and Software*, vol. 30, no. 1, pp. 94–127, 2015.

[24] A. Goldberg and R. Tarjan, "Finding minimum-cost circulations by canceling negative cycles," in *20th ACM Symposium on Theory of Computing*, Jan. 1988.

[25] OpenStreetMap contributors, "OpenStreetMap," 2015, www.openstreetmap.org/copyright.

[26] M. M. Grynbaum, "Where do all the cabs go in the late afternoon?" *The New York Times*, Jan. 2011.