

TOP: Vehicle Trajectory based Driving Speed Optimization Strategy for Travel Time Minimization and Road Congestion Avoidance

Li Yan and Haiying Shen

Department of Electrical and Computer Engineering

Clemson University, Clemson, SC 29631

Email: {lyan, shenh}@clemson.edu

Abstract—Traffic congestion control is pivotal for intelligent transportation systems. Previous works optimize vehicle speed for different objectives such as minimizing fuel consumption and minimizing travel time. However, they overlook the possible congestion generation in the future (e.g., in 5mins), which may degrade the performance of achieving the objectives. In this paper, we propose a vehicle Trajectory based driving speed Optimization strategy (*TOP*) to minimize vehicle travel time and meanwhile avoid generating congestion. Its basic idea is to adjust vehicles' mobility to alleviate road congestion globally. *TOP* has a framework for collecting vehicles' information to a central server, which calculates the parameters depicting the future road condition (e.g., driving time, vehicle density, and probability of accident). The server then formulates a non-cooperative Stackelberg game considering these parameters, in which when each vehicle aims to minimize its travel time, the road congestion is also proactively avoided. After the Stackelberg equilibrium is reached, the optimal driving speed for each vehicle and the expected vehicle density that maximizes the utilization of the road network are determined. Our real trace analysis confirms some characteristics of vehicle mobility to support the design of *TOP*. Extensive trace-driven experiments show the effectiveness and superior performance of *TOP* in comparison with other driving speed optimization methods.

I. INTRODUCTION

In recent decades, Intelligent Transportation Systems (ITSs) have received much attention. The ITSs summarize advanced applications aiming at providing innovative services related to different modes of transportation and traffic management. To support the operation of various ITS applications, traffic congestion control is very important for urban road networks [1]–[3] when trying to maximize their utilization. For example, the road management authorities hope that the density of vehicles simultaneously passing through each road is lower than a threshold so that the overall road network keeps operable. Also, the public transit service vehicles require their covered routes to be non-congested so that they can follow their schedule on time. However, due to the high mobility of vehicles and difficulty in controlling vehicle speeds, congestion control in urban road networks is a non-trivial task.

In recent years, many methods have been proposed to reduce vehicles' travel time by adaptively controlling traffic signal [4], [5] or suggesting optimal speeds to vehicles for different objectives such as minimizing fuel

consumption and travel time [6]–[10]. In the former group of methods [4], [5], the controller at a road intersection properly schedules the passing of vehicles to minimize the vehicles' total travel time caused by red lights or long queues. In the latter group of methods [6]–[10], the optimal driving speed of a vehicle is determined based on the vehicle's real-time driving information (e.g., fuel consumption, traffic state). However, these methods overlook the possible road congestion generation in the future (e.g., in 5mins), which may degrade the performance of achieving the objectives. In other words, these methods cannot avoid the generation of road congestion globally in the road network in the future. By “in the future”, we mean in a future time during a vehicle's driving time period. For example, before “rush hours”, arterial roads may be non-congested. However, if legions of vehicles drive by the currently “optimal speeds” in their individual routes, they may crowd into the arterial roads simultaneously, which results in congestion.

However, solving this neglected problem is non-trivial. The road congestion is measured by vehicle density; a higher vehicle density increases the utilization of the road network but generates congestion and decreases vehicle speed, and vice versa. Therefore, it is a challenge to maximize the utilization of the road network while proactively avoiding congestion and maximizing the vehicle speed. In this paper, we aim to tackle this challenge by proposing a vehicle Trajectory based driving speed Optimization strategy (*TOP*) that uses game theory to let vehicle drive as fast and safely as possible, and meanwhile proactively avoid generating road congestion in the future. Its basic idea is to periodically adjust vehicles' mobility to alleviate road congestion globally. The vehicles report their information to a central server through road-side-units (RSUs) located alongside the roads. The central server calculates each vehicle's trajectory in the next time slot (denoted by \mathcal{T}_{c+1}) and determines the parameters depicting the future utilization of the road network (e.g., vehicle density, driving time and probability of accident). This is based on the previous observation that vehicles' trajectories can soundly illustrate the future mobility of the vehicles [11]–[15]. To maximize the utilization of the road network while minimizing the probability of road congestion, the central server formulates a non-cooperative Stackelberg game, in which each vehicle aims at minimizing its travel time and maximizing safety

while avoiding generating congestion in the future. After the Stackelberg equilibrium is reached, when vehicles follow their optimal speeds, it also proactively avoids generating congestion in the future. Moreover, the road network can be fully utilized without imbalanced or high vehicle density (i.e., congestion) in road segments. In summary, our contributions include:

- (1) Our analysis on two real vehicle traces [16], [17] confirms the characteristics of vehicle mobility and lays the foundation for the design of *TOP*.
- (2) We propose a non-cooperative Stackelberg game based vehicle speed optimization strategy to find the optimal speed for each vehicle that enables it to drive as fast and safely as possible, while avoiding the generation of congestion in the future.
- (3) We have conducted extensive trace-driven experiments to show the effectiveness of *TOP* in maximizing the utilization of road network, avoiding congestions, and satisfying drivers' need of driving as fast and safely as possible.

Within our knowledge, this work is the first to provide vehicle speed optimization that aims at letting vehicles drive as fast and safely as possible, while proactively avoiding the generation of road congestion in the future. The remainder of the paper is organized as follows. Section II presents related works. Section III presents the trace analysis and findings that support *TOP*. Section IV presents the detailed design of *TOP*. Section V presents trace-driven experimental results. Section VI concludes the paper and marks future research direction.

II. RELATED WORK

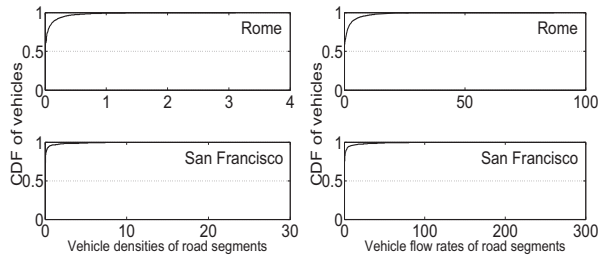
Real-time traffic based vehicle speed optimization. Several methods for vehicle speed optimization with different objectives have been proposed. Kouvelas *et al.* [4] proposed a hybrid approach for traffic signal control considering the saturation status of the road. Pandit *et al.* [5] proposed a vehicular network based method that collects and aggregates real-time traffic information to optimize signal control. Tseng *et al.* [10] proposed a vehicle density estimation scheme using neighbor tables communicated between vehicles. Chen *et al.* [9] proposed to use VANETs to send queries between source and destination back and forth, and selects the path with the shortest time. Ozatay *et al.* [7] used cloud computing [18], [19] in optimizing vehicle speed profile by solving a dynamic programming problem. Asadi and Vahidi [8] proposed a control algorithm to enable vehicles to approach traffic light at green as much as possible, thereby saving fuel and reducing travel time. Groot *et al.* [6] proposed to model vehicle-congestion relationship as reverse Stackelberg games to optimally distribute traffic over road network, and meanwhile ensure that each vehicle can finish travel within its expected travel time. For the vehicles with the same origin-destination, the central server uses different pricing of freeways (e.g., longer route has lower price) to induce these vehicles to

choose different routes to distribute the traffic. However, this method overlooks that the vehicles with different origin-destination pairs may compete for the same road segment. Also, it does not aim to minimize the travel time of the vehicles. As indicated previously, the above methods do not consider whether the currently suggested speed will cause congestion to certain road segments in the future. To solve this problem, *TOP* firstly utilizes vehicles' trajectories to extract the parameters of future road traffic, and then uses the parameters in formulating a non-cooperative Stackelberg game that aims to let vehicles drive as fast and safely as possible, and meanwhile avoid the generation of congestion in future.

Vehicle future mobility based routing. Many works [20], [21] focus on using vehicles' current or historical mobility statistics to predict the vehicles' future mobility. Some other works [11]–[15] found that utilizing vehicles' GPS trajectory to deduce the vehicles' future mobility is reliable for data delivery in vehicular networks. Wu *et al.* [11] found the spatio-temporal correlation in vehicle mobility and noted that the future trajectory of a vehicle is correlated with its past trajectory. In Trajectory-based Data Forwarding Scheme (TBD) [12], Trajectory-based Statistical Forwarding Scheme (TSF) [13], [14] and Shared-Trajectory-based Data Forwarding Scheme (STDFS) [15], trajectory information of vehicles is collected through access points and used to predict the vehicle mobility for data forwarding. Our work is based on the observations in these works that vehicles' trajectories can soundly illustrate the future mobility of vehicles, which is used to estimate road vehicle density in the future.

III. TRACE ANALYSIS

In this section, we present our trace analysis on the Rome trace [16] and the San Francisco trace [17], which demonstrates the characteristic of vehicle mobility in urban area and provides the rationale of the design of *TOP*. Both of the traces are 30-day taxi traces. Since taxis move persistently and cover almost the entire road networks, their movement can illustrate the traffic state [2], [22]–[24]. In these traces, each taxi reports location record (timestamp, ID, GPS position) to a central server every 7 seconds. We filtered out the positions with precision larger than 20 meters and taxis with few appearances (<500), and extracted road layout from vehicle movement. The position records of vehicles are highly overlapped, so we extracted intersections where vehicles make turns as landmarks and simplified vehicles' movement records to sequences of landmarks. Finally, Rome has 315 taxis and 4638 landmarks, and San Francisco has 536 taxis and 2508 landmarks. When a vehicle stays at one position for more than 5mins, we call this position *anchor position* and consider it as the ending position of the previous trajectory and the starting position of the next trajectory. Thus, the anchor positions separate each vehicle's trace into several trajectories.



(a) Vehicle densities. (b) Vehicle flow rates.

Fig. 1: Concurrent competition for road usage.

A. Concepts and Problem Introduction

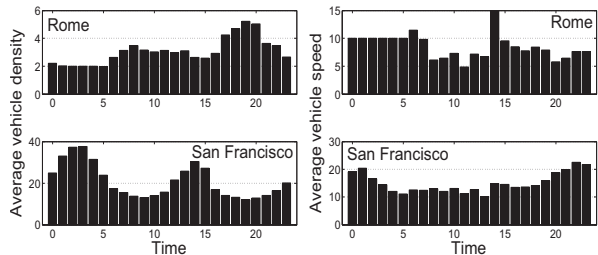
We define a road segment (denoted by s_i) as the road link between two neighboring intersections (i.e., landmarks). Vehicle density of road segment s_i (denoted by d_i) is defined as the average number of vehicles per mile in the road segment ($veh/mile$), and the flow rate of road segment s_i (denoted by f_i) is defined as the average number of vehicles driving through the segment per unit time [6], [25] (veh/h). The vehicle flow rate of segment s_i equals to the product of vehicle density and average vehicle speed on s_i (denoted by v_i) [26];

$$f_i = d_i \cdot v_i. \quad (1)$$

The road congestion is measured by vehicle density, and the utilization of the road network is measure by flow rate. Therefore, in order to increase the utilization of the road segment s_i (i.e., f_i), we need to increase vehicle density (d_i) and/or vehicle speed (v_i). However, higher vehicle density may lead to congestion and hence lower vehicle speed. Therefore, it is a challenge to maximize the utilization of road network and meanwhile maximize vehicle speed, which is the objective of this paper.

B. Concurrent Competition for Road Usage

Previous methods locally control traffic or compute suggested speed based on current traffic state on each vehicle's scheduled route. If a vehicle follows the speed individually optimized for it, due to the ignorance of other vehicles' mobility, some arterial roads may become crowded with many vehicles, that is, the vehicles concurrently compete for these roads. To confirm this conjecture, we measured the cumulative distribution function (CDF) of the vehicle density and the CDF of the flow rate on all road segments as shown in Figure 1(a) and Figure 1(b). We calculated the vehicle density and vehicle flow rate for 30 days and draw their average values. The vehicle density and vehicle flow rate are sampled every 30mins on every road segment per day. We see that for the Rome trace, the vehicle density of 90% of road segments is less than $0.5veh/mile$, and the vehicle flow rate of 90% of road segments is less than $10veh/h$. But the other 10% of road segments have vehicle density and vehicle flow rate as high as $3veh/mile$ and $60veh/h$, respectively. For the San Francisco trace, the vehicle density of 95% of road segments is less than $3veh/mile$, and the vehicle flow rate of 95% of road segments is less than $25veh/h$. But the other 5%



(a) Vehicle densities over time. (b) Vehicle speeds over time.

Fig. 2: Vehicles' temporal preference on roads.

of road segments have vehicle density and vehicle flow rate as high as $24veh/mile$ and $50veh/h$, respectively. These results demonstrate that in the urban road network, vehicles usually concurrently compete for usage on few popular roads, resulting in their excessive utilization. Therefore, we can try to distribute traffic evenly in the road network, i.e., achieve similar vehicle density in all road segment, in order to avoid the congestion and increase the utilization of road network. The cause to repeated congestion on arterial roads is excessive concurrent utilization of vehicles [6]. Therefore, we further analyze vehicles' temporal preference on driving roads.

C. Vehicles' Temporal Preference on Roads

If the vehicle density on a road exceeds a threshold, the driving speed of vehicles on the road is likely to be affected due to congestion. This is especially true for arterial roads since they are quite likely to be over-utilized during rush hours. To verify such intuition, we measured the average vehicle density and average vehicle speed on the most highly utilized road segment of the two traces (road segment 4433 in the Rome trace, road segment 0 in the San Francisco trace) hourly during each day in the 30 days, which are shown in Figure 2(a) and Figure 2(b), respectively. We see that for the Rome trace, the vehicle densities during $6:00\sim 13:00$ and $17:00\sim 20:00$ are higher than the other hours. In contrast, the average vehicle speeds during these two periods are lower than the other hours. For the San Francisco trace, the vehicle densities during $1:00\sim 4:00$ and $12:00\sim 15:00$ are higher than the other hours. In contrast, the average vehicle speeds during these two periods are generally lower than the other hours. These results demonstrate that excessively high vehicle density deteriorates road driving condition, which causes reduced driving speed. The results confirm that avoiding congestion is important to increasing driving speed and reducing travel time, especially in rush hours.

IV. SYSTEM DESIGN

A. Overview

ITSs support the installation of RSUs alongside road segments to provide communication between vehicles and the central server [14], [27], [28]. As shown in Figure 3, we establish a three-layer information collection and dissemination framework, which consists of

vehicles as the service layer, RSUs as the communication backbone and a central server as the computation layer. Each vehicle contacts the central server through RSUs.

As in the traffic management papers in [29], we consider road segments have equal vehicle density limits. In this paper, we focus on optimizing the vehicles' speed on their original route. We leave the optimal route selection as future incremental work. To let vehicles drive as fast and safely as possible while avoiding generating congestion on the road network, we use the Stackelberg

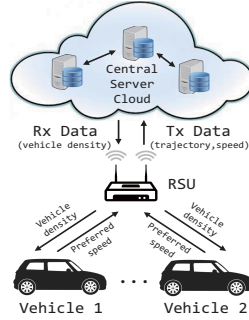


Fig. 3: System structure.

game [30] between the vehicles and the central server to determine the expected vehicle density that maximizes the utilization of the road network and optimal driving speed for each vehicle. The gaming process is executed periodically with period \mathcal{T} (e.g., 5mins) as follows:

- 1) Through a nearby RSU, the vehicle reports its current position and intended destination to the central server.
- 2) Based on the information collected from vehicles, the central server calculates the trajectory of each vehicle in \mathcal{T}_{c+1} and predicts the vehicle density in each road segment at the next time slot. Then, a gaming process is conducted between the central server and each vehicle.
- 3) Based on the predicted average vehicle density per road segment in the road network in \mathcal{T}_{c+1} , the central server determines a set of expected average densities that are achievable by vehicle speed adjustment.
- 4) Based on each expected average density, each vehicle determines its speed that maximizes its utility (speed and safety) and reports the speed to the central server.
- 5) The central server determines the final expected average density that maximizes its utility (maximizing flow rate of the road network) and notifies all vehicles.
- 6) Each vehicle chooses its speed corresponding to the final expected average density.

With the optimal speeds, the vehicle density of each road segment will be approximately the determined vehicle density. Thus, the total traffic in the road network is well balanced with no congestion and its utilization is maximized. We will first explain how the central server predicts the vehicle density of road segments (Section IV-B) and then present the non-cooperative Stackelberg gaming (Section IV-C).

B. Future Road Vehicle Density Prediction

The gaming process runs after each time slot \mathcal{T} (e.g., 5mins). For example, when the central server starts the game at 00:00, it needs to estimate the vehicle density of each road segment in [00:00,00:05] in order to determine an achievable vehicle density in the entire road network for vehicles to choose their optimal speeds.

In this section, we present how to estimate the vehicle density of each road segment in \mathcal{T}_{c+1} with current vehicle speeds. First, the central server needs to determine each vehicle's trajectory in \mathcal{T}_{c+1} . It consists of the road segments it will pass in \mathcal{T}_{c+1} and their corresponding travel times $\{(s_i, \tilde{T}_i) | i = 1, 2, \dots, M\}$, where s_i denotes the i th road segment, \tilde{T}_i denotes the estimated travel time from current position to s_i and M denotes the number of road segments that the vehicle will pass in \mathcal{T}_{c+1} .

Then, by modeling the arrival times as normal random variables, *TOP* sums up the vehicles' probabilities of appearance on each road segment as its vehicle density at the next time slot. The average vehicle density per road segment will be used in the driving speed optimization gaming presented in Section IV-C. After each vehicle determines its speed in gaming, the vehicle density will be updated and used for the next gaming process.

1) *Trajectory Calculation*: A vehicle periodically reports its current position and its destination to the central server. To generate the vehicle's trajectory in \mathcal{T}_{c+1} , the central server first determines the sequence of road segments connecting the current position and the destination based on road topology [31]. It then calculates the travel time of each road segment that will be passed in \mathcal{T}_{c+1} by the vehicle. After a gaming process, a vehicle's optimal speed on s_i is determined, denoted by v_i . Then, for each road segment s_i , the estimated travel time on s_i (denoted by \tilde{t}_i) can be calculated by

$$\tilde{t}_i = l_i / v_i, \quad (2)$$

where l_i is the length of s_i . A problem is how to estimate the travel time of s_i initially when no game has been played. To handle this problem, we use the current vehicle density of the road segment to estimate the speed for the vehicle as in traditional vehicle density based speed estimation methods [26]. It has been indicated that for a road segment s_i , its reachable speed is related to a vehicle density limit d_i^m . When the vehicle density is below d_i^m , vehicles on the road segment can drive with the free flow speed (i.e., speed limit, denoted by v_i^{max}). If the vehicle density exceeds d_i^m , the road segment will be congested and the vehicles have to drive with the congested speed (denoted by v_i^{min}). d_i^{jam} is the vehicle density that will cause s_i to be completely jammed. d_i^m can be obtained from field observation, and d_i^{jam} can be obtained from the road network's designed capacity [26]. Currently, the vehicle density of each road segment can be well monitored [2], [22]–[24], [32]. Then, we can roughly estimate the allowed vehicle speed under current vehicle density for each road segment as below:

$$\tilde{t}_i = \begin{cases} l_i / v_i^{max}, & 0 \leq d_i < d_i^m \\ l_i / v_i^{min}, & d_i^m \leq d_i < d_i^{jam} \\ \infty, & d_i \geq d_i^{jam} \end{cases} \quad (3)$$

The trajectories generated by GPS do not consider the road congestion condition and hence may not be sufficiently accurate. In contrast, *TOP* calculates the trajectories of vehicles considering future road congestion.

The travel times of a road segment can be described by normally distributed and statistically independent random variables with acceptable precision [33], [34]. Therefore, the estimated travel time from the current position to road segment s_i is the sum of the travel times of the composing road segments from current position to s_i , $\tilde{T}_i = \sum_{m=1}^{M_i} \tilde{t}_m$, where M_i is the number of road segments from current position to s_i . When $\tilde{T}_i \geq \mathcal{T}$, the trajectory for \mathcal{T}_{c+1} has been generated. Based on the historical data of the estimated travel time of s_m and real travel time on s_m of all vehicles, the central server can calculate variance σ_m^2 . Then, the standard deviation of \tilde{T}_i is calculated by summing the variances of the composing road segments, $\Delta_i^2 = \sum_{m=1}^{M_i} \sigma_m^2$.

2) *Road Vehicle Density Calculation*: The estimated travel time in $\{(s_i, \tilde{T}_i) | i = 1, 2, \dots, M\}$ only has a certain probability to be accurate. Then, we have two steps to calculate the vehicle density of each road segment in \mathcal{T}_{c+1} . First, we use a vehicle's trajectory in \mathcal{T}_{c+1} to estimate the probability that the vehicle will appear at each road segment in its trajectory in \mathcal{T}_{c+1} . Then, we calculate the sum of all the vehicles' appearance probabilities at a road segment in \mathcal{T}_{c+1} as the vehicle density of the road segment in \mathcal{T}_{c+1} .

Suppose the next time slot is the j th time slot in a day, represented by $\mathcal{T}_{c+1} = [t_j^s, t_j^e]$ (e.g., [00:00,00:05]), where t_j^s and t_j^e are the starting time and ending time of the time slot, respectively. For each vehicle, *TOP* uses its estimated travel time to s_i to measure its appearance probability at s_i during $[t_j^s, t_j^e]$. Therefore, the vehicle's appearance probability at s_i during $[t_j^s, t_j^e]$ is

$$P(T_i \leq t_j^e - t_j^s) = \Phi\left(\frac{t_j^e - t_j^s - \tilde{T}_i}{\Delta_i}\right) - \Phi\left(\frac{-\tilde{T}_i}{\Delta_i}\right) \quad (4)$$

where T_i denotes the vehicle's actual travel time from current position to s_i , and $\Phi(\cdot)$ is the CDF of the standard normal distribution with mean \tilde{T}_i and standard deviation Δ_i . The CDF for each vehicle on each road segment s_i is calculated based on the historical records of all vehicles' travel times on the road segment. Then, for each road segment s_i , the central server estimates its vehicle density in \mathcal{T}_{c+1} by summing up the appearance probabilities of vehicles (P_k) on s_i during \mathcal{T}_{c+1} :

$$d_{i+1}^{s_i} = \sum_{k=1}^N P_k(T_i \leq t_j^e - t_j^s) \quad (5)$$

where N is the number of vehicles that will pass s_i during $[t_j^s, t_j^e]$. For example, given current time 00:00, the estimated vehicle density of *College Ave* for \mathcal{T}_{c+1} , namely 00:00~00:05, is 26.16 vehicles/mile.

3) *Safety Estimation*: Each road segment has a probability of accident occurrence. The probability depends on the structure feature of the road segment (e.g., the degree of straightness, sharp turn, road surface bump) and the traffic condition. It has been verified that traffic conditions (e.g., heavy traffic volume, speeding) affect the likelihood of accident occurrence [35]. The traffic condition of a road segment has a long-term pattern,

that is, the vehicle flow rate at each time slot remains similar irrespective of days. For example, people are likely to encounter congestion on their way to work during morning rush hours every workday.

TOP relies on historical records of accidents to depict the likelihood of accident for each road segment in each time interval in a day [36], [37]. Considering that the probability of accident is time-varying (e.g., some road segments are more likely to have accidents in Winter than in Spring), *TOP* uses a time window to control the number of days for consideration. The larger the window size is, the more accident events that can be captured. Specifically, the accident probability of road segment s_i during j th time interval $[t_j^s, t_j^e]$ is calculated by:

$$p_i^j = \frac{\sum_{w=1}^W T_j^w}{W \cdot (t_j^e - t_j^s)} \quad (6)$$

where p_i^j is the accident probability of s_i during the j th time interval, W is the window size, and T_j^w is the length of time that s_i is affected by accident during the j th time interval in the w th day. Finally, for each road segment, the central server generates a table summarizing accident probability during each time interval, as shown in Table I. Since higher vehicle density leads to shorter distance between vehicles, which renders higher risks of accident, we relate p_i^j with vehicle density in determining the utility of drivers in gaming in Section IV-C.

TABLE I: Table of accident probability of road segment *College Ave*.

Time	Accident probability
00:00~00:05	0.05
00:05~00:10	0.02
...	...

C. Driving Speed Optimization Gaming

1) *Overview*: On one hand, traveling quickly (i.e., short driving time and no congestion) and safely (i.e., no accident) is desired by drivers. On the other hand, the transportation authority hopes to maximize the utilization of the road network (i.e., maximum vehicle flow rate). Based on Equation (1), to increase road network utilization, we need to increase the vehicle density, which however may lead to road congestion. Then, the vehicle speed drops down and results in low flow rate and hence low utilization of the road network.

It is found that drivers may drive slower given a higher specified vehicle density in order to keep a safe inter-vehicle distance to keep safety, especially for the drivers with high probability of accidents [35]. Therefore, the drivers will adjust speeds in response to a given vehicle density. Thus, we can formulate the speed optimization as a non-cooperative Stackelberg game [6], [38] between the central server and the drivers, where the central server is the leader and the drivers are followers.

In the Stackelberg game, the leader considers the predicted average vehicle density of a road segment (introduced in Section IV-B2), and then chooses a set of expected vehicle densities (denoted by $D = \{d_1, d_2, \dots, d_n\}$) that are achievable by vehicle speed adjustment. The

central server hopes to evenly distribute the vehicles over all road segments by properly assigning a d value. The followers receive D from the leader and picks a speed in response to each d_i to maximize its own utility (driving as fast and safely as possible while minimizing the risk of congestion). Next, the central server selects the vehicle density that maximizes its utility (i.e., vehicle flow rate of the road network), denoted by d_l and then the vehicles choose their speeds corresponding to the selected d_l . Finally, we solve the Stackelberg equilibrium of the game, i.e., the game reaches a state that the road network utilization is maximized while the drivers are satisfied with the driving status (judged by driving speed and associated risk of congestion). The gaming is executed periodically. In the following, we first introduce the utility of a driver and the utility of the central server, and then introduce the gaming between them.

2) *Utility Function of Drivers:* For drivers, we define a utility function to quantify the level of benefit that a driver obtains from driving by a speed on road segment s_i . It is calculated by subtracting the potential risk of congestion ($U_r(\cdot)$) from a vehicle's satisfaction degree ($U_s(\cdot)$), as shown in Equation (8).

$$\begin{aligned} F(v_i, \alpha_i, p_i^j) &= U_s(v_i, \alpha_i, p_i^j) - U_r(d, v_i, p_i^j) \quad (7) \\ \text{s.t. } v_i &\leq v_i^{max} \end{aligned}$$

where v_i is the vehicle's speed for optimization; α_i is a scale factor to make $U_s(\cdot)$ and $U_r(\cdot)$ comparable; p_i^j is calculated by Equation (6).

$U_s(\cdot)$ ought to be non-decreasing as each driver desires high speed (i.e., short driving time). Meanwhile, the marginal satisfaction degree of the driver is non-increasing because the driver's satisfaction degree gradually gets saturated when the vehicle speed increases to some level [26]. Moreover, $U_s(\cdot)$ is inversely related with the probability of accident because a lower possibility of accident corresponds to higher level of satisfaction [39]. Considering these properties, we design $U_s(\cdot)$ as a concave function. Since the Natural Logarithmic Functions are representative concave functions [40], we define:

$$U_s(v_i, \alpha_i, p_i^j) = \alpha_i \cdot \ln(v_i + p_i^{j-1}) \quad (8)$$

A driver's potential risk of congestion is determined by the accident probability of the road segment (p_i^j) and vehicle flow rate [35] (Equation (1)).

$$U_r(d, v_i, p_i^j) = p_i^j dv_i \quad (9)$$

The utility of a driver decreases with a higher vehicle density and vice versa. Combining Equation (8) and Equation (9) into Equation (8), we have:

$$\begin{aligned} F(v_i, \alpha_i, p_i^j) &= \alpha_i \cdot \ln(v_i + p_i^{j-1}) - p_i^j dv_i \quad (10) \\ \text{s.t. } v_i &\leq v_i^{max} \end{aligned}$$

Note the gaming is executed periodically, so it is possible that a vehicle may enter other road segments during the current time slot. We use γ_i to denote the percentage of \mathcal{T} that the vehicle will spend on segment s_i . Then, the utility of the vehicle is calculated by:

$$\begin{aligned} \sum_i \gamma_i F(v_i, \alpha_i, p_i^j) \quad (11) \\ \text{s.t. } v_i \leq v_i^{max} \end{aligned}$$

3) *Utility Function of Central Server:* The central server always aims at maximizing the vehicle flow rate on overall road network:

$$L(d) = \sum_{i=1}^{N_s} d_i \cdot v_i \quad (12)$$

where N_s is the total number of road segments.

4) *Optimal Driving Speed Selection:* Recall that based on Equation (5), the central server predicts the vehicle densities of all road segments. It then calculates the average estimated vehicle density of the road network during next period of gaming: $\overline{d_{c+1}} = \sum_{k=1}^{N_s} d_{i+1}^{s_k} / N_s$. Based on $\overline{d_{c+1}}$, the central server determines a range of expected vehicle densities that are achievable by vehicle speed adjustment, and offers these densities to each vehicle for selection, which is defined as:

$$d_u = \ln(u + 1) \cdot \overline{d_{c+1}}, u \in [1, \dots, n] \quad (13)$$

We use $D = \{d_1, d_2, \dots, d_n\}$ to denote the n levels of expected vehicle densities for \mathcal{T}_{c+1} . In practice, n should be at least larger than the exponent so that the vehicle has multiple selections around $\overline{d_{c+1}}$. The central server notifies drivers of the D . If $\overline{d_{c+1}}$ leads to an increased expected vehicle density (d_u), the drivers will be encouraged to decrease driving speed in order to drive safely. Otherwise, the drivers will be encouraged to increase driving speed in order to increase benefit while maintaining driving safety. Note the increment rate of $U_s(\cdot)$ (Natural Logarithmic Function) is slower than $U_r(\cdot)$ (Linear Function) when speed v_i increases. Therefore, according to Equation (8), increasing driving speed on current road segment (v_i) will reduce a driver's utility because $U_r(\cdot)$ will increase faster than $U_s(\cdot)$. Thus, driving at a slower speed can prevent the vehicle density of the road network from further increasing, i.e., prevent traffic congestion.

For each $d_u \in D$, if a driver will drive in its current road segment s_i during the next time slot, it chooses a new speed that maximizes its utility $F(\cdot)$, denoted by v_{iu} , as shown in Equation (14).

$$v_{iu} = \arg \max_{v_i \leq v_i^{max}} F(v_i, \alpha_i, p_i^j) \quad (14)$$

If a driver will drive through more than one road segment s_i, s_j, \dots , it chooses a set of speeds in each of the segments to maximize its utility $F(\cdot)$, denoted by $\{v_{iu}, v_{ju}, \dots\}$ as shown in Equation (15).

$$\{v_{iu}, v_{ju}, \dots\} = \arg \max_{v_k \leq v_k^{max}} \sum_k \gamma_k F(v_k, \alpha_k, p_k^j) \quad (15)$$

Finally, the driver reports the n candidate speeds to the central server. The central server finalizes the expected vehicle density (d_l) that maximizes its utility $L(\cdot)$ based on the candidate speeds from all drivers.

$$d_l = \arg \max_{d_u \in D} L(d_u) = \arg \max_{d_u \in D} d_u \sum_{N_s} v_{iu} \quad (16)$$

The central server then notifies all drivers of the new expected vehicle density d_l . Then, among the n candidate speeds, each driver picks the speed corresponding to d_l .

V. PERFORMANCE EVALUATION

A. Experimental Settings

We conducted trace-driven experiments based on the Rome [16] and the San Francisco [17] traces introduced in Section III. Unless otherwise specified, the experiment settings are the same as those in Section III. The number of accidents occurred in Rome and San Francisco in each month are obtained from [41] and [42], respectively. The window size W was set to 7 days and $T_j^w = 1h$. The scale factor α_i was set to 2.85 for Rome and 5 for San Francisco. We measure a driver's satisfaction degree when (s)he drives on road segment s_i with speed v_i by $\ln(v_i + p_i^{j-1}) / \ln(v_i^{max} + p_i^{j-1})$ (deduced from Equation (8)). The gaming procedures are launched every 15mins in the two traces. To simulate that vehicles drive by their optimal speeds, we dynamically update the timestamps of arrivals at landmarks according to the vehicles' optimal speeds. Therefore, in the experiment, the vehicles follow the movement paths recorded in the traces but with modified timestamps.

We compared *TOP* with the traffic signal control method [5] (*Signal* in short) and the vehicle speed optimization method [7] (*RealSpeed* in short). *Signal* uses vehicular ad hoc networks (VANETs) to formulate vehicles into platoons. The controller at each intersection uses the oldest-arrival-first scheduling algorithm to arrange the passing of platoons so that the vehicles' total travel time is minimized. In *RealSpeed*, by aiming at reducing fuel consumption and satisfying driver with reduced travel time, the vehicle speed is optimized by dynamic programming constrained by speed limit, real-time traffic and driver's destination. To make *RealSpeed* comparable to the other methods, we excluded its fuel consumption constraint in our experiments. *Signal* and *RealSpeed* cannot proactively avoid the generation of road congestion in the future. Each experiment is for 30 days. In each hour throughout each day, we measured the following metrics and report the average value in each hour for the 30 days.

- *Average vehicle speed*: The average speed of all the speeds determined by the games during an hour.
- *Average flow rate*: After each game, we calculate the flow rate per road segment by $\sum_{i=1}^{N_s} d_i \cdot v_i / N_s$. Then, we calculate the average flow rate per road segment in all the games during an hour.
- *Average driving time*: The average driving time on each road segment for all the travels on segments during an hour.
- *Average driver satisfaction*: The average satisfaction degree of the drivers after travels per hour.

B. Experimental Results

1) Average Vehicle Speed:

Figure 4(a) and Figure 5(a) show the average speed of vehicles during different time intervals with the Rome and San Francisco traces, respectively. We see that for Rome, the average vehicle speeds follow: *TOP* > *RealSpeed* > *Signal*. While for San Francisco, the average vehicle speeds follow: *TOP* > *Signal* > *RealSpeed*.

TOP always has much higher average vehicle speed than others in both traces. Before optimization, the future traffic on the scheduled route has been deduced by the central server from the vehicle trajectories. Although the results might have deviation from the true results, they are still effective in predicting the future movement of vehicles. With gaming using the predicted vehicle density, *TOP* enables the central server to maximally avoid road congestion caused by confluent vehicle flows. Meanwhile, each vehicle can drive by a speed as fast and safely as possible. As a result, *TOP* generates the highest average speed of vehicles. Both *RealSpeed* and *Signal* do not proactively avoid generating congestions in the future and congestions decrease vehicle speeds, thus producing lower average speed of vehicles.

RealSpeed has the secondary performance in Rome, but the lowest performance in San Francisco. In *RealSpeed*, for a vehicle, the server generates a route based on the collected information of the intended travel. Then, the server collects the associated traffic and geographical information, and calculates optimal vehicle speed aiming at reducing travel time through dynamic programming. However, since San Francisco has many uniformly distributed road segments with short lengths [17], the vehicle flow on the vehicle's scheduled route can be easily congested by vehicle flows from other intersected road segments. In contrast, the road segments in Rome have fewer intersections [16]. Therefore, the vehicle traffic in the road network is less likely to be congested than that in San Francisco.

Signal has the lowest average vehicle speed in Rome, but the second lowest performance in San Francisco. This is because *Signal* aims at reducing vehicles' travel time near intersections rather than in the global road network. In *Signal*, the vehicle flows on the road network are partitioned into several platoons of vehicles. By viewing each platoon as a job, the traffic management problem is formulated as a job scheduling problem at intersections. To minimize the time of vehicles passing the intersection, *Signal* utilizes the oldest-arrival-first scheduling algorithm. However, Rome's road segments are quite crowded at popular sites and have short distance [16], which make streets near popular sites heavily utilized. Locally minimizing the time at certain intersections inevitably exacerbates congestion at the other intersections. Therefore, *Signal* cannot achieve an optimal solution in the whole road network in Rome. In contrast, the road segment distribution of San Francisco is more uniform than that in Rome [17]. Therefore,

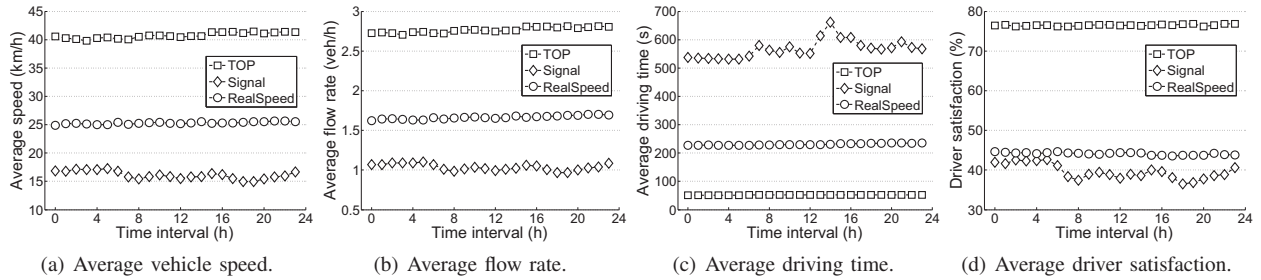


Fig. 4: Performance over time with the Rome trace.

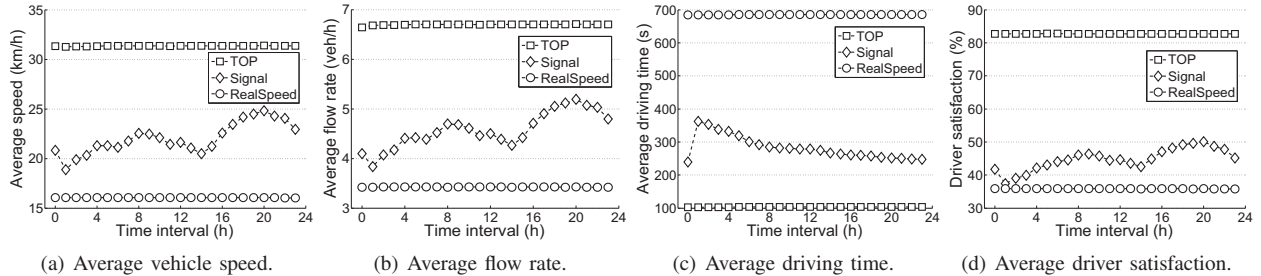


Fig. 5: Performance over time with the San Francisco trace.

Signal can better schedule vehicles passing through the intersections in San Francisco than in Rome, resulting in the better performance of *Signal* in San Francisco.

2) Average Flow Rate:

Figure 4(b) and Figure 5(b) show the average flow rate during different time intervals with the Rome and San Francisco traces, respectively. We see that for Rome, the average vehicle flow rates follow: $TOP > RealSpeed > Signal$. While for San Francisco, the results follow: $TOP > Signal > RealSpeed$.

The average flow vehicle rates follow the same trend as the average vehicle speed result. Higher speed means that the vehicle flow can move faster on road segment as long as the vehicle density does not result in congestion. Although some road segments may be too crowded to let vehicles maintain high speeds, their flow rate is still large as long as their vehicle density does not exceed the jam level. Through comparing Figure 4(a) and Figure 5(a) with Figure 4(b) and Figure 5(b), we can see that although the average speed keeps above $15 km/h$, the vehicle flow rate is as low as $1 veh/h$. This shows that when the road network is non-congested, the vehicles in *Signal* and *RealSpeed* can drive as fast as possible (i.e., speed limit), which results in acceptable average driving speed. While without proactively avoiding congestion, the vehicle flows may generate congestion.

3) Average Driving Time:

Figure 4(c) and Figure 5(c) show the average driving time during different time intervals with the Rome and San Francisco traces, respectively. We see that for Rome, the average vehicle driving time follows: $Signal > RealSpeed > TOP$. While for San Francisco, the average vehicle driving time follows: $RealSpeed > Signal > TOP$.

TOP always has the lowest driving time because each vehicle can drive by a fast speed with low probability of suffering from congestion. *Signal* has the highest driving

time in Rome, and the second highest driving time in San Francisco. Correspondingly, *RealSpeed* has the second highest driving time in Rome, and the highest driving time in San Francisco. These results are consistent with those of the average vehicle speed due to the same reasons. It is noticeable that in San Francisco, there is a heap between $0h$ and $1h$. This is because there is a drop of speed during this time interval. When multiple vehicles simultaneously enter an intersection, but traffic signals cannot schedule their passing in time, the vehicles then wait in queues at the intersection.

4) Average Driver Satisfaction:

Figure 4(d) and Figure 5(d) show the average driver satisfaction during different time intervals with the Rome and San Francisco traces, respectively. We see that for Rome, the average satisfaction follows: $TOP > RealSpeed > Signal$. While for San Francisco, the average satisfaction follows: $TOP > Signal > RealSpeed$.

Driver satisfaction is jointly determined by vehicle speed and accident probability. Since the accident probability is calculated offline and does not change during vehicles' movement, drivers' satisfaction is solely determined by the vehicle speed. Since *TOP* generates the highest vehicle speed, it always ranks the highest and achieves over 80% satisfaction in both traces. The satisfaction results are consistent with the average vehicle speed results due to the same reasons.

VI. CONCLUSION

Previous works for speed optimization does not proactively avoid the generation of congestion in the future. We proposed *TOP*, a vehicle trajectory based driving speed optimization strategy aiming at minimizing each vehicle's travel time while avoiding generation of congestion. Our analysis on the vehicle mobility and congestion based on two real-world traces support the motivation for the design of *TOP*. *TOP* uses

vehicle trajectories to estimate the vehicle density of each road segment in the near future. Then, by using a non-cooperative Stackelberg game between each vehicle and the central server, the vehicle's driving speed is optimized so that it can drive as fast and safely as possible while proactively avoiding generating congestion. We have conducted extensive experiments based on the two traces. The experiment results validate the high effectiveness of *TOP* and its superior performance compared to previous methods in terms of the utilization of road network, congestions, and driver satisfaction. In our future work, we plan to consider vehicles' social relationship in avoiding road congestion and develop more reasonable schemes to motivate vehicle cooperation.

ACKNOWLEDGEMENTS

This research was supported in part by U.S. NSF grants NSF-1404981, IIS-1354123, CNS-1254006, and Microsoft Research Faculty Fellowship 8300751.

REFERENCES

- [1] S. Lin, B. De Schutter, Y. Xi, and H. Hellendoorn, "Integrated urban traffic control for the reduction of travel delays and emissions," *IEEE TITS*, vol. 14, no. 4, 2013.
- [2] R. Herring, A. Hofleitner, P. Abbeel, and A. Bayen, "Estimating arterial traffic conditions using sparse probe data," in *Proc. of ITSC*, 2010.
- [3] L. Li, D. Wen, and D. Yao, "A survey of traffic control with vehicular communications," *IEEE TITS*, vol. 15, no. 1, 2014.
- [4] A. Kouvelas, K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "A hybrid strategy for real-time traffic signal control of urban road networks," *IEEE TITS*, vol. 12, no. 3, 2011.
- [5] K. Pandit, D. Ghosal, H. M. Zhang, and C.-N. Chuah, "Adaptive traffic signal control with vehicular ad hoc networks," *IEEE TVT*, vol. 62, no. 4, 2013.
- [6] N. Groot, B. De Schutter, and H. Hellendoorn, "Toward system-optimal routing in traffic networks: a reverse stackelberg game approach," *IEEE TITS*, vol. 16, no. 1, 2015.
- [7] E. Ozatay, S. Onori, J. Wollaeger, U. Ozguner, G. Rizzoni, D. Filev, J. Michelini, and S. Di Cairano, "Cloud-based velocity profile optimization for everyday driving: A dynamic-programming-based solution," *IEEE TITS*, vol. 15, no. 6, 2014.
- [8] B. Asadi and A. Vahidi, "Predictive cruise control: Utilizing upcoming traffic signal information for improving fuel economy and reducing trip time," *IEEE TCST*, vol. 19, no. 3, 2011.
- [9] W. Chen, S. Zhu, and D. Li, "VAN: Vehicle-assisted shortest-time path navigation," in *Proc. of MASS*, 2010.
- [10] Y.-T. Tseng, R.-H. Jan, C. Chen, C.-F. Wang, and H.-H. Li, "A vehicle-density-based forwarding scheme for emergency message broadcasts in vanets," in *Proc. of MASS*, 2010.
- [11] Y. Wu, Y. Zhu, and B. Li, "Trajectory improves data delivery in vehicular networks," in *Proc. of INFOCOM*, 2011.
- [12] J. Jeong, S. Guo, Y. Gu, T. He, and D. H. Du, "Trajectory-based data forwarding for light-traffic vehicular ad hoc networks," *IEEE TPDS*, vol. 22, no. 5, 2011.
- [13] J. Jeong, S. Guo, Y. Gu, T. He, and D. H. Du, "Trajectory-based statistical forwarding for multihop infrastructure-to-vehicle data delivery," *IEEE TMC*, vol. 11, no. 10, 2012.
- [14] J. Jeong, S. Guo, Y. Gu, T. He, and D. H. Du, "TSF: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks," in *Proc. of ICDCS*, 2010.
- [15] F. Xu, S. Guo, J. Jeong, Y. Gu, Q. Cao, M. Liu, and T. He, "Utilizing shared vehicle trajectories for data forwarding in vehicular networks," in *Proc. of INFOCOM*, 2011.
- [16] R. Amici, M. Bonola, L. Bracciale, P. Loreti, A. Rabuffi, and G. Bianchi, "Performance assessment of an epidemic protocol in VANET using real traces," in *Proc. of MoWNeT*, 2014.
- [17] M. Piórkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "A parsimonious model of mobile partitioned networks with clustering," in *Proc. of COMSNETS*, 2009.
- [18] L. Chen and H. Shen, "Consolidating complementary vms with spatial/temporal-awareness in cloud datacenters," in *Proc. of INFOCOM*, 2014.
- [19] C. Qiu, H. Shen, and L. Chen, "Towards green cloud computing: Demand allocation and pricing policies for cloud service brokerage," in *Proc. of Big Data*, 2015.
- [20] C. Lochert, B. Scheuermann, C. Wewetzer, A. Luebke, and M. Mauve, "Data aggregation and roadside unit placement for a vanet traffic information system," in *Proc. of VAINET workshop*, 2008.
- [21] T. Nadeem, S. Dashtinezhad, C. Liao, and L. Iftode, "Trafficview: traffic data dissemination using car-to-car communication," *Proc. of SIGMOBILE*, vol. 8, no. 3, 2004.
- [22] Tao, Sha and Manolopoulos, Vasileios and Rodriguez, Saul and Rusu, Ana, "Real-time urban traffic state estimation with A-GPS mobile phones as probes," *JTT*, vol. 2, no. 01, 2012.
- [23] Q.-J. Kong, Q. Zhao, C. Wei, and Y. Liu, "Efficient traffic state estimation for large-scale urban road networks," *IEEE TITS*, vol. 14, no. 1, 2013.
- [24] C. Furtlehner, J.-M. Lasgouttes, and A. de La Fortelle, "A belief propagation approach to traffic prediction using probe vehicles," in *Proc. of ITSC*, 2007.
- [25] F. Bai and B. Krishnamachari, "Spatio-temporal variations of vehicle traffic in VANETs: facts and implications," in *Proc. of VANET*, 2009.
- [26] Y. Xu, Q.-J. Kong, S. Lin, and Y. Liu, "Urban traffic flow prediction based on road network model," in *Proc. of ICNSC*, 2012.
- [27] K. C. Dey, L. Yan, X. Wang, Y. Wang, H. Shen, M. Chowdhury, L. Yu, C. Qiu, and V. Soundararaj, "A review of communication, driver characteristics, and controls aspects of cooperative adaptive cruise control (cacc)," *IEEE TITS*, vol. 17, no. 2, 2016.
- [28] Y. Lin and H. Shen, "VShare: A wireless social network aided vehicle sharing system using hierarchical cloud architecture," in *Proc. of IoTDI*, 2016.
- [29] E. Brockfeld, R. Barlovic, A. Schadschneider, and M. Schreckenberg, "Optimizing traffic lights in a cellular automaton model for city traffic," *Physical Review E*, vol. 64, no. 5, 2001.
- [30] H. Shen and Z. Li, "New bandwidth sharing and pricing policies to achieve a win-win situation for cloud provider and tenants," in *Proc. of INFOCOM*, 2014.
- [31] B. Wu, H. Shen, and K. Chen, "Exploiting active sub-areas for multi-copy routing in VDTNs," in *Proc. of ICCCN*, 2015.
- [32] A. Sarker, C. Qiu, H. Shen, A. Gil, J. Taiber, M. Chowdhury, J. Martin, M. Devine, and A. Rindos, "An efficient wireless power transfer system to balance the state of charge of electric vehicles," in *Proc. of ICPP*, 2016.
- [33] R. Li, G. Rose, and M. Sarvi, "Using automatic vehicle identification data to gain insight into travel time variability and its causes," *JTRB*, no. 1945, 2006.
- [34] R. Li, H. Chai, and J. Tang, "Empirical study of travel time estimation and reliability," *Mathematical Problems in Engineering*, vol. 2013, 2013.
- [35] M. A. Abdel-Aty and A. E. Radwan, "Modeling traffic accident occurrence and involvement," *Accident Analysis & Prevention*, 2000.
- [36] L. Yan, H. Shen, and K. Chen, "TSearch: Target-oriented low-delay node searching in dtns with social network properties," in *Proc. of INFOCOM*, 2015.
- [37] J. Liu, L. Yu, H. Shen, Y. He, and J. Hallstrom, "Characterizing data deliverability of greedy routing in wireless sensor networks," in *Proc. of SECON*, 2015.
- [38] K. Chen, H. Shen, and L. Yan, "Multicent: A multifunctional incentive scheme adaptive to diverse performance objectives for dtn routing," *IEEE TPDS*, vol. 26, no. 6, 2015.
- [39] E. Freitas, C. Mendonça, J. A. Santos, C. Murteira, and J. Ferreira, "Traffic noise abatement: How different pavements, vehicle speeds and traffic densities affect annoyance levels," *Transportation Research*, vol. 17, no. 4, 2012.
- [40] K. G. Binmore, *Mathematical Analysis: a straightforward approach*. Cambridge University Press, 1982.
- [41] "Rome accident statistics," <http://www.telegraph.co.uk>, accessed July 29, 2015.
- [42] "San Francisco accident statistics," <http://www.city-data.com>, accessed July 29, 2015.