

FLEET-ORIENTED REAL-TIME VEHICULAR TRACKING AT URBAN SCALE

BY XIAOYANG XIE

**A thesis submitted to the
Graduate School—New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Master of Science
Graduate Program in Computer Science**

**Written under the direction of
Desheng Zhang
and approved by**

New Brunswick, New Jersey

May, 2017

ABSTRACT OF THE THESIS

Fleet-oriented Real-Time Vehicular Tracking at Urban Scale

by Xiaoyang Xie

Thesis Director: Desheng Zhang

Nowadays, vehicular sensing has become increasingly important to collect urban data to understand and address mobility challenges. A straightforward method to achieve this goal is to build fine-grained city-scale sensing infrastructures to instrument all vehicles with sensors and centralized communication interfaces, which leads to very expensive costs. Therefore, previous work in urban sensing explores less expensive methods in two categories: (i) Centralized methods where a small number of well-equipped vehicles with centralized real-time cellular connections to upload sensing data in real time, which leads to data sparsity due to limited number of vehicles; (ii) Distributed methods where a large number of minimally- equipped vehicles with peer-to-peer communication devices to upload sensing data in an offline fashion, which leads to long delay due to peer-to-peer communication.

To address these issues, this dissertation explores a new direction of combining centralized sensing and distributed sensing together for a hybrid vehicular sensing framework based on two new opportunities, as a part of intelligent transportation system (i) Recently, we have witnessed a surge of commercial vehicular fleets, e.g., taxis, buses, and trucks, with advanced sensing, and centralized/distributed communication devices. (ii) There has been a trend to consider mandating all private vehicles to broadcast their

status (potential sensing data) to nearby vehicles, e.g., using peer-to-peer communications to broadcast safety message to nearby vehicles including ID, speed, locations, and sensing status for safety applications. Therefore, the key question this thesis answers is can we use a small number of well-equipped commercial vehicles to track (and then collect data from) a large number of minimally- equipped private vehicles for urban scale sensing in real time.

Real-time vehicle tracking at urban scale is essential to various urban services. To track vehicles at individual levels, most existing approaches rely on static infrastructures (e.g., cameras) or mobile services (e.g., smartphone apps). However, these approaches are often inadequate for urban-scale individual tracking because of their static natures or low penetration rates. In this thesis, we design a tracking system called coTrack to utilize commercial vehicular fleets (e.g., taxis, buses, and trucks) for real-time vehicle tracking at urban scale, given (i) the availability of well-equipped commercial fleets, and (ii) an increasing trend of mandating all vehicles to broadcast their status for safety applications. The key technical challenge we addressed is how to recover spatiotemporal tracking gaps by considering various mobility patterns of commercial vehicles with a hidden Markov model. We evaluate coTrack with a preliminary road test and a large-scale trace-driven evaluation based on vehicular fleets in the Chinese city Shenzhen, including 14 thousand taxis, 13 thousand buses, 13 thousand trucks, and 10 thousand private vehicles. We compare coTrack to infrastructure and cellphone-based approaches, and the results show that we increase the tracking accuracy by 42.2% and 23.2% on average. Further, we design a service to utilize private vehicle tracking to infer travel time between $491 * 491$ region pairs, and the results show that given the diverse mobility of private vehicles, we can infer travel time between 15% more region pairs than using commercial vehicles alone.

Acknowledgements

I wish to thank my advisor, Professor Desheng Zhang, who provides so many invaluable guidance, advice, efforts, and suggestions in my thesis. I have received numerous treasures from Professor Zhang, such as his knowledge, vision consideration, and inspiration. I also want to thank the members of my committee, Professor Zheng Zhang, Professor Jingjin Yu and Professor Richard P Martin for their generosity in taking the time to review and provide insightful suggestions to improve this dissertation significantly. This dissertation was only possible through the contributions from them. I appreciate their efforts. I would also like to thank the many other Rutgers Computer Science faculty and staff members who taught and help me all the way throughout my master time at Rutgers.

Dedication

To my father, Yulong Xie, and my mother, Youhua Zhuang

Table of Contents

Abstract	ii
Acknowledgements	iv
Dedication	v
1. Introduction	1
1.1. Thesis	1
1.2. Background	1
1.3. Related Work	4
1.4. Motivations	6
2. Three-Layer Architecture	9
3. Physical Infrastructure Layer	11
4. Mobility Modeling Layer: Design	14
4.1. Challenge: Spatiotemporal Gaps	14
4.2. HMM-based Modeling	15
4.3. Collaborative Tracking	18
4.3.1. Key Idea	18
4.3.2. Mobility Pattern Prediction	18
4.3.3. Online Prediction Updating	21
4.4. Summary	21
5. Mobility Modeling Layer: Field Study	22
6. Mobility Modeling Layer: Evaluation	24
6.1. Evaluation Data Management	24

6.2. Evaluation Methodology	24
6.3. Evaluation Results	27
6.3.1. Baseline Comparison	27
6.3.2. Impacts of Fleet Types	28
6.3.3. Impacts of Locations and Time	29
6.3.4. Impact of Tracking Parameters	30
6.4. Evaluation Summary	31
7. Application Layer	33
7.1. Application Background	33
7.2. Application Evaluation	34
8. Discussion	37
9. Conclusion	39
References	41

Chapter 1

Introduction

1.1 Thesis

The thesis of the dissertation states that:

By using a hybrid sensing framework, a small number of well-equipped commercial vehicles can be used to track and collect data from urban-scale minimally-equipped private vehicles for urban sensing in real time based on their mobility patterns.

1.2 Background

Nowadays, vehicles are essential components for our daily life, e.g., we have 1.2 billion vehicles in the world by 2015, and this number is projected to increase to 2 billion by 2035 [13]. This phenomenon is more obvious in urban areas, e.g., in New York City, there are 2.9 million vehicles entering the city each day [1], and in Beijing, there are 5.6 million vehicles on the road [14]. All these vehicles in urban areas lead to various challenges, e.g., traffic congestion and energy consumption [44]. To address these challenges, it is essential to understand mobility patterns of these vehicles, i.e., tracking urban-scale vehicles in real time.

However, tracking urban-scale vehicles in real time is extremely challenging due to fine temporal coverage (e.g., 10 second intervals), large spatial coverage (e.g., all road segments), and high quantitative coverage (e.g., all vehicles), which requires a major investment of infrastructures. The existing approaches for vehicle tracking are mostly based on (i) static infrastructures, e.g., cameras [12] and RFID [8]; (ii) mobile services,

e.g., vehicular manufacturers’ services (e.g., OnStar [19], Ford Sync [35], and BMW Assist [3]), smartphone apps (e.g., Google Maps [11] and Apple Maps [2]), and navigators (e.g., Garmin [10]). However, the static infrastructure-based systems can only track vehicles in limited locations with pre-deployed infrastructures, e.g., intersections with cameras; the mobile service based systems can only track limited vehicles due to their penetration rates at urban scale [37] [38]. Thus, they cannot enable urban-scale vehicle tracking in real time.

Recently, two new opportunities emerge based on urban infrastructure upgrades, which have the potential to enable urban-scale vehicle tracking in real time. (i) We have been witnessing a surge of commercial fleets [5] instrumented with sensing and communication capabilities, e.g., cellular connection and dedicated short range communications (DSRC), enabling real-time sensing and data uploading [7]. (ii) There has been a trend to consider mandating all vehicles to broadcast their status to nearby vehicles for safety applications [9] [4], e.g., using DSRC to broadcast basic safety message to nearby vehicles within 100 to 300 meters at a frequency of 2 to 10 times per second including ID, speed, and locations [21]. Based on these two opportunities, the key question we are trying to answer is that “can we utilize small-scale yet well-equipped commercial vehicles to accurately track large-scale yet minimally-equipped private vehicles with existing urban infrastructures?”

In this dissertation, we answered this question by designing an urban-scale system called coTrack for collaborative fleet-oriented vehicle tracking and resultant applications. The core idea of coTrack is to (i) collect private vehicles’ location data by various commercial vehicles through distributed sensing and communications (e.g., DSRC-based broadcasting), and then (ii) consolidate these collected data on the cloud by real-time centralized communications (e.g., cellular-based uploading) to infer the detailed traces of private vehicles. Different from the existing approaches (i.e., static infrastructures [12] or mobile services [37] [38]), coTrack utilizes a *mobile infrastructure* approach based on existing commercial fleets potentially without additional investments. This is because (i) local broadcasting of private vehicles is most likely to become mandatory in the near future for safety applications [6], and (ii) centralized

status uploading for commercial vehicles has already been mandatory for accounting in many cities [15]. The key technical challenge in coTrack is how to fuse heterogeneous commercial fleet data to recover urban-scale private vehicles’ traces in real time based on (i) diverse mobility patterns of fleets (i.e., random taxis, semi-random trucks, and regular buses), and (ii) contextual information (e.g., road map, traffic speeds, and historical trips). In particular, the key contributions of the dissertation are as follows:

- To our knowledge, we conduct the first urban-scale vehicle tracking based on heterogeneous fleets. Our work advances the state-of-the-art vehicular investigation in two aspects: (i) the most comprehensive vehicular systems, including taxis, buses, trucks, and private vehicles from the same city, and (ii) detailed GPS traces from 50 thousand vehicles, more than 3% of all vehicles in the studied city. Our infrastructures and data are at least one or two orders of magnitude larger than existing academic systems (e.g., GreenGPS [23], EasyTracker [18], VTrack [38], and CTrack [37]).
- We present a three-layer system called coTrack for collaborative vehicle tracking and its resultant applications at urban-scale in real time: (i) a physical infrastructure layer for data collection from heterogeneous urban fleets including taxis, buses, and trucks; (ii) a mobility modeling layer for individual-based private vehicle tracking at urban scale in real time; (iii) an application layer for private-vehicle trace-driven services. coTrack has a modularized architecture where the mobility modeling layer separates upper-layer services from lower-layer urban infrastructures.
- We design an inference technique for the mobility modeling layer based on hidden Markov models to infer real-time traces of private vehicles as the core of our modeling layer. It utilizes diverse data from commercial vehicles to collaboratively infer locations of private vehicles at fine-grained spatiotemporal partitions. The tracking results are further improved by real-world contextual constraints including detailed road map, real-time traffic speeds, and historical trips.
- We implement and evaluate coTrack in Shenzhen with a preliminary road test

and a trace-driven evaluation based on real-world data from 14 thousand taxis, 13 thousand buses, 13 thousand trucks, and 10 thousand private vehicles. To our knowledge, coTrack is one of the largest urban vehicular systems driven by real-world fleets and their data. We evaluate coTrack by comparing it to infrastructure and cellphone based approaches, and the results show that we increase the tracking accuracy by 42.2% and 23.2% on average.

- Based on the inferred private vehicles’ traces, we design a service at the application layer to estimate the urban-scale travel time between 491×491 region pairs in real time. Thanks to diverse mobility patterns of private vehicles, we obtain interregional travel time between 15% more region pairs than using commercial vehicles alone, showing coTrack’s real-world value.

1.3 Related Work

Real-time vehicle tracking at urban scale is crucial for real-world applications, e.g., navigation, traffic control, and location-based services [44]. Basically, all the existing systems can be divided into four categories.

- **Static City Infrastructures:** Public infrastructures, e.g., cameras and RFID-based toll stations, are widely used in cities for traffic monitoring, crime monitoring and fee charging, which can be used potentially for urban-scale vehicle tracking [44]. In New York City, there are more than 643 closed circuit television cameras for real-time traffic cameras [12], and more than 88 thruways supporting E-ZPass, which is an RFID-based charging system [8]. Wireless access points as a part of urban infrastructures can be used to detect positions of vehicles [31], which thus can be potentially used for tracking. However, all these systems suffer from the sparsity and static natures of infrastructures. In contrast, our coTrack system is based on urban fleets to track vehicles in a mobile fashion.
- **Manufacturer Services:** Many vehicle manufacturers have services to track their own vehicles for different applications, e.g., navigation and maintenance,

e.g., OnStar [19], Ford Sync [35] and BMW Assist [3]. Typically, these systems obtain locations of vehicles in real time by built-in GPS and communicate with servers through cellular connections. However, these systems are only available in specific brands and have very limited urban vehicle coverage.

- **Smartphone-based Systems:** Smartphone-based systems are an important research area for vehicle tracking because of its various sensors [30]. Several systems are proposed to track vehicles in real time [28] [43] [47] [37] [20], to monitor traffic conditions [38] [46], to find real-time parking spots [33], to infer transportation modality [36], and to predict bus arrival times [18]. Generally, these systems rely on smartphones and utilize one or multiple sensors to detect locations and movements of vehicles. However, these systems are limited by low penetration rates of apps and are hard to use for urban-scale vehicle tracking.
- **Fleet-based Systems:** Several systems based on large-scale urban fleet data have also been proposed, e.g., inferring real-world road maps [17]; estimating city traffic volumes for drivers [15]; querying the expected duration and fare of a planned taxi trip [16] [26]; predicting passenger demand for taxi drivers [25]; recommending optimal pickup locations [24] [32]; modeling the urban transit [45]; detecting the taxi anomaly [34]; navigating new drivers based on GPS traces of experienced drivers [41] [40]. However, these systems are typically used for one particular fleet, e.g., taxis or buses, and are not focused on the private vehicle tracking.

Based on our analyses, almost all the above approaches are limited for urban-scale vehicle tracking by infrastructure coverage, low penetration rates, deployment scale, or system objectives. These limitations motivate us to take advantages of commercial fleets that are already deployed in the city with data collection capability to track vehicles at urban scale. coTrack utilizes a mobile infrastructure approach combining heterogeneous fleets with a mobility-driven approach, which makes our work significantly different from the state-of-the-art approaches.

1.4 Motivations

In this section, we show our design motivations by investigating opportunities and challenges for commercial fleet-oriented vehicle tracking based on real-world urban fleet GPS data in the Chinese city Shenzhen. The details of fleet systems and their data are introduced in Section 3.

Opportunities for Fleet-oriented Tracking: The spatiotemporal road coverage of commercial fleets indicates their capability for vehicle tracking at urban scale in real time. Based on GPS data from three commercial fleets (i.e., a 14-thousand taxi fleet, a 13-thousand bus fleet, and a 13-thousand freight truck fleet), we show the percentage of 160 thousand road segments with at least one vehicle from a particular fleet in one-minute slots in Figure 1.1.

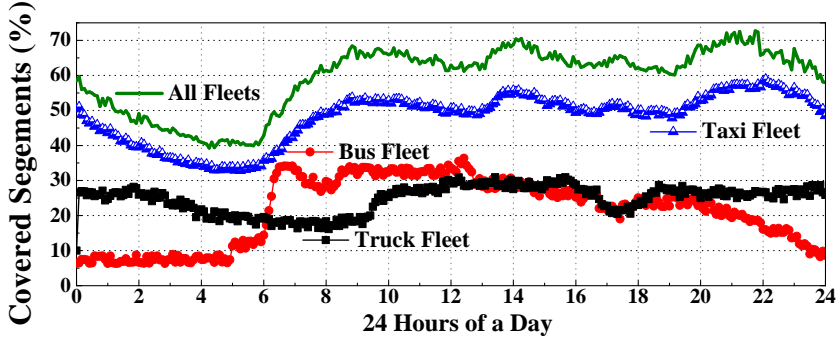


Fig 1.1: Spatiotemporal Road Coverage

We found that these fleets, if combining together, cover a high percentage of segments under one-minute slots, e.g., 61% of segments on average, which indicates commercial fleets have a fine-spatiotemporal coverage. Thus they have a high potential to track private vehicles at urban scale.

Challenges for Fleet-oriented Tracking: Along with GPS data from the above commercial fleets, we utilize GPS data from 10 thousand private vehicles to show the challenges of fleet-oriented tracking based on their real-time mobility patterns. We envision that if a private vehicle is within 100 meters of a commercial vehicle for more than 10 seconds, it can be tracked by this commercial vehicle (e.g., based on local broadcasting through DSRC [7]). We will discuss these parameter settings and details

of private vehicles data in Section 3. We show the percentage of private vehicles that do not have any commercial vehicle within 100 meters for more than 10 seconds (i.e., untrackable vehicles) during one-minute slots on 24 hours of a day in Figure 1.2.

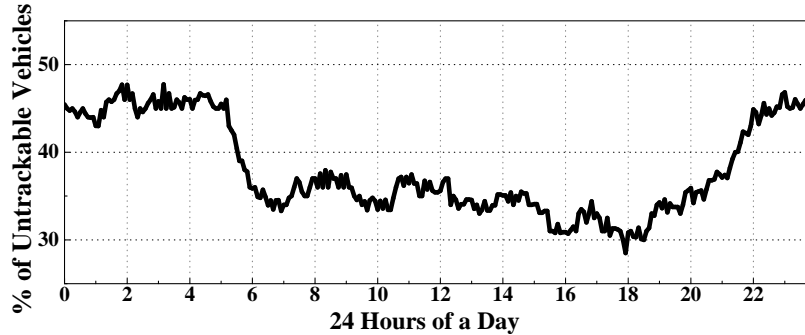


Fig 1.2: Untrackable Vehicles

We found that in the daytime from 6 AM to 8 PM, more than 35% of the private vehicles in our dataset are untrackable, i.e., they do not have any commercial vehicles in its proximity of 100 meters longer than 10 seconds in one-minute slots. During the early morning and late night, this percentage increases to 45%. This is because commercial vehicles cannot cover all spatiotemporal areas of a city due to the random mobility and limited quantity, leading to spatiotemporal tracking gaps. It indicates utilizing commercial fleets to track private vehicles, though promising as shown in Figure 1.1, still has a key challenge to address, i.e., inferring detailed traces from all these private vehicles with spatiotemporal tracking gaps. To understand these spatiotemporal tracking gaps, we calculate the percentages for the time of all vehicle's trips without any commercial vehicle in its 100-meter radius longer than 10 seconds and cluster all these trips by their start time in Figure 1.3.

We found that for vehicles' trips starting from daytime, they cannot be tracked by any commercial vehicles during more than 30% of their total time. During the early morning and late night, it increases to 45%, which is a significant gap if we want to track these vehicles in real time.

Summary: We explore the opportunities for urban-scale vehicle tracking based on

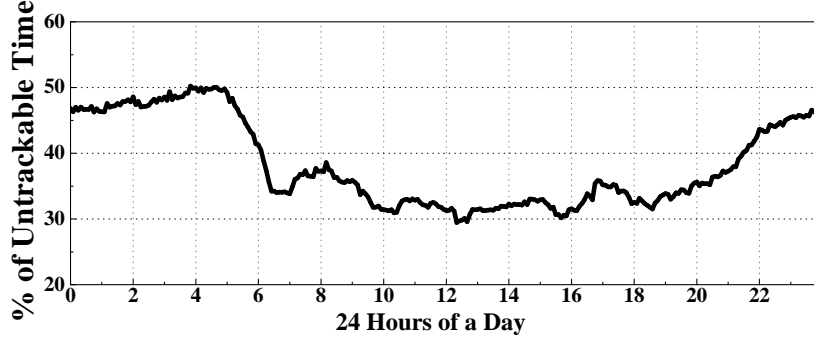


Fig 1.3: Spatiotemporal Tracking Gaps

urban commercial fleets given their spatiotemporal road coverage. However, we also identify major spatiotemporal tracking gaps for fleet-oriented tracking in term of untrackable vehicles and untrackable time. These gaps have to be addressed in order to enable urban-scale vehicle tracking by commercial fleets in real time. It motivates us to design a system to collaboratively utilize commercial fleets to track private vehicles as follows.

In the rest of this dissertation, we organize the dissertation as follows. §2 presents a system overview. §3 depicts the physical infrastructure layer. §4, §5 and §6 give the design, implementation, and evaluation of the mobility modeling payer. §7 describes a real-world service at the application layer, followed by the discussion in §8. §9 concludes the dissertation.

Chapter 2

Three-Layer Architecture

In the coTrack system, we conceptually consider a set of heterogeneous urban fleets (e.g., taxi, bus, and truck fleets) as a virtual mobile sensor network to track private vehicles at urban scale in real time. Built upon an integration of various large-scale commercial fleets, coTrack provides unseen mobility dynamics for individual vehicles from a mobile infrastructure perspective under extremely fine-grained spatiotemporal resolutions to support real-world services. In general, these services cannot be achieved by either static infrastructures or mobile services, e.g., OnStar [19], Google Maps [11], or academic systems [37] [38].

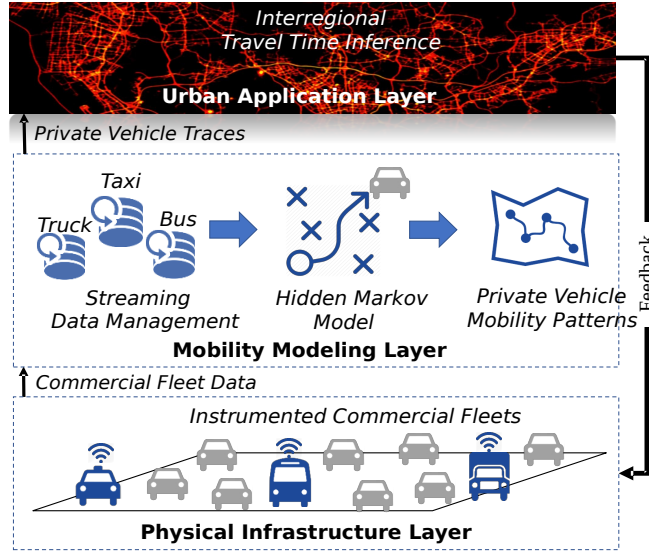


Fig 2.1: coTrack Architecture

Figure 2.1, we outline the coTrack architecture with three layers, *i.e.*, Physical Infrastructure Layer, Mobility Modeling Layer, and Urban Application Layer. These three

layers span the whole coTrack data-processing chain. We provide a roadmap for the rest of the dissertation as in Figure 2.1. (i) In §3, we introduce the physical infrastructure layer where we collect commercial fleet data from three large-scale fleets. (ii) In §4, §5 and §6, we present the design, implementation, and evaluation for our modeling layer based on a hidden Markov model (HMM) to infer real-time traces of private vehicles based on periodically-uploaded data from commercial vehicles and real-time contextual information. (iii) In §7, to close the control loop, we design and evaluate a service in our application layer to estimate interregional travel time based on recovered real-time traces of private vehicles. We envision that drivers would use this service to find efficient routes, which, in return, provides positive feedback to urban fleets. As a result, with a highlight on collaborative fleet-oriented tracking, coTrack builds an architectural bridge between small-scale yet well equipped commercial vehicles and large-scale yet minimally equipped private vehicles to enable novel services.

Chapter 3

Physical Infrastructure Layer

Taxi Fleet		Bus Fleet		Truck Fleet		Private Vehicles	
Beginning	1/1/2012	Beginning	1/1/2013	Beginning	12/1/2013	Beginning	2/1/2016
# of Taxis	14,453	# of Buses	13,032	# of Trucks	45,356	# of Vehicles	293,849
Size	3.5 TB	Size	1.2 TB	Size	1.1 TB	Size	600 GB
# of Records	29 billion	# of Records	11 billion	# of Records	9 billion	# of Records	0.8 billion
Format		Format		Format		Format	
Plate ID	Date&Time	Plate ID	Date&Time	Plate ID	Date&Time	Device ID	Date&Time
Status	GPS&Speed	Stop ID	GPS&Speed	GPS	Speed	GPS	Speed

Fig 3.1: Fleets and Their Data

We have been collaborating with several service providers and the Shenzhen Committee of Transportation (SCT) for real-time fleet access. As in Figure 3.1, we consider three commercial fleets, i.e., taxi, bus, and truck fleets, in this version of implementation, which detect private vehicles from complimentary perspectives.

- Taxi Fleet:** We access the Shenzhen taxi fleet and their data through SCT to which all taxi companies upload their taxi status in real time through a centralized connection, i.e., cellular networks, with monthly fees. The taxi fleet in Shenzhen has 14 thousand taxis generating one status record per 30 seconds including GPS locations, time, speed, etc. The taxi fleet has a random mobility pattern to cover most of the road segments in Shenzhen as we show in Figure 1.1.
- Bus Fleet:** We access the Shenzhen bus fleet including 976 bus lines and their bus data through SCT to which all buses upload their status in real time by cellular networks. The Shenzhen bus fleet has 13 thousand buses, and their status records are generated every 30 seconds when buses are operating. Compared to the taxi fleet, the bus fleet has a regular pattern due to their operating routes. As a result, their spatial patterns are fixed, while their temporal patterns are varied because of real-time traffics even with a fixed timetable.

- **Truck Fleet:** We access a truck fleet with 45 thousand trucks, among which 13 thousand trucks are operating in Shenzhen, by working with a large logistics company. In general, every truck uploads its status records including GPS locations and travel speeds back to a company server every 15 seconds on average for real-time monitoring, which then are routed to our server in real time. Most of these trucks are for delivery, and a truck typically has an urban area to cover, but its daily delivery schedule changes based on actual demand, leading to a semi-random mobility pattern.

The above urban fleet access enables urban-scale phenomenon modeling in real time. In coTrack, we mainly focus on private vehicle tracking and evaluation.

- **Distributed Private Vehicle Detection:** Currently, in our fleet platform in Shenzhen, most of the vehicles (both commercial and private vehicles) are not equipped with sensing and peer-to-peer communication devices, e.g., DSRC. By working with a taxi company, we equipped 106 taxis with sensing and communication devices (details in § 5), but for now we cannot have an urban-scale prototype system for a large-scale tracking. In this project, given the recent trend of mandating DSRC at many countries [9], we envision a scenario that many vehicles are equipped with peer-to-peer communication devices (e.g., DSRC), and then a commercial vehicle can detect a private vehicle if they are within the communication range. Under this scenario, we evaluate our idea of utilizing commercial vehicles to track private vehicles.
- **Real-world Private Vehicle Data:** As part of our physical infrastructures, we have access to a private vehicle network with more than 293 thousand private vehicles, among which 10 thousand vehicles are in Shenzhen. The data of these private vehicles are collected by onboard navigators from a large tech company in Shenzhen. When a navigator is turned on, a GPS record is uploaded every 10 seconds to a server of this company and then is routed to our server. These GPS records are mainly used for navigation services. With such fine-grained data access, we can design and validate our coTrack system accordingly.

Our endeavor of accessing such heterogeneous large-scale commercial fleets and consolidating their data enables extremely large-scale fine-grained urban mobility study as in Figure 3.1. Comparing to the most state-of-the-art systems [23] [18] [38] [37], our infrastructure is unprecedented in terms of the fleet size and complementary mobility patterns. For example, Figure 3.2 gives a heatmap visualization of these four fleets based on their one-day data. We found that each fleet has its own unique mobility pattern shown by the circles, e.g., (i) the taxi fleet covers most urban areas; (ii) the truck fleet mostly is focused on highways and a few industrial areas; (iii) the bus fleet is focused on major road segments; (iv) the private fleet has similar patterns with taxis but with some exceptions at a few residential areas. As follows, we introduce our modeling layer to track private vehicles with these commercial fleets in §4.

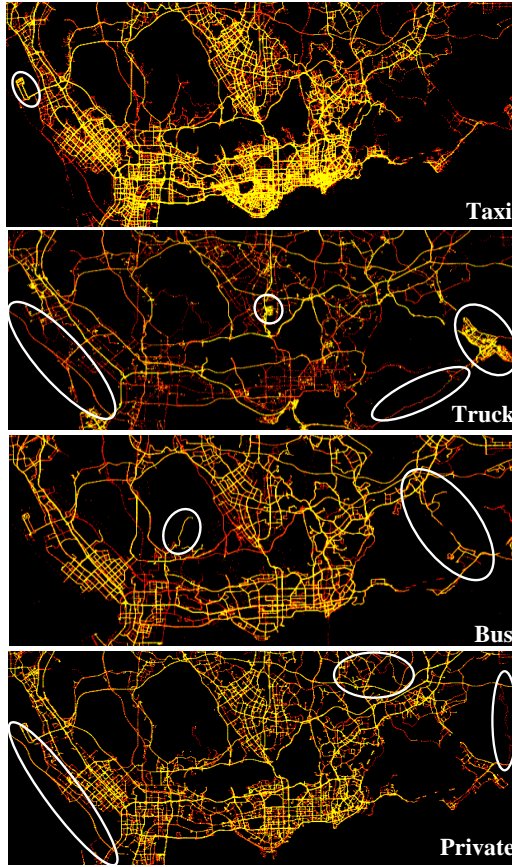


Fig 3.2: Fleet Visualization

Chapter 4

Mobility Modeling Layer: Design

Based on commercial vehicles' data, we track private vehicles in real time. We first present a deeper data-driven analysis for the key challenge, i.e., spatiotemporal gaps, and then present the core design of vehicle tracking by a hidden Markov model (HMM), and finally give a summary.

4.1 Challenge: Spatiotemporal Gaps

Since a commercial vehicle can receive data from a private vehicle including real-time locations of this private vehicles, a naive solution would be to simply use these locations to track private vehicles. However, such a solution does not work because a private vehicle does not always have commercial vehicles in its communication range for a period of time. To support our claim, we study a joint mobility pattern of commercial and private vehicles based on our real-world fleet data in Shenzhen. Our model is based on the assumption that a commercial vehicle can track a private vehicle if they are closer than 100 meters for more than 10 seconds. Note that we set 100 meters as a peer-to-peer communication range and 10 seconds as a contact duration given urban environments with interferences, although in practice some peer-to-peer communication devices, e.g., DSRC, have a range of 300-500 meters in open spaces and shorter contact durations [21]. We evaluate these parameters § 6. In this dissertation, we define a time period during which a private vehicle does not have any commercial vehicle in its communication range as a spatiotemporal gap. A longer spatiotemporal gap potentially leads to a lower accuracy of our modeling. We perform a discretization process where we divide time into 10-second slots to study mobile interactions between private and commercial vehicles. As follows, we study the frequency and duration of these gaps

using 10-second slots in Figures 4.1 and 4.2.

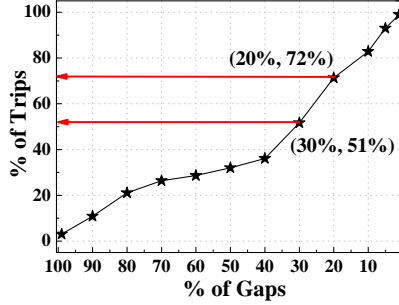


Fig 4.1: Gap Frequency

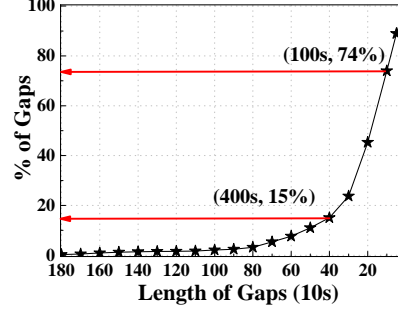


Fig 4.2: Gap Duration

Figure 4.1 gives a distribution of spatiotemporal gaps based on our one-week GPS data of private and commercial vehicles including taxi, bus, and truck fleets. We found that 51% of the private vehicle trips have at least 30% of travel time during which there are no commercial vehicles in its communication range; 72% of the private vehicle trips have at least 20% of travel time without any commercial vehicles nearby. As a result, a naive solution where we directly use private vehicles' locations from commercial vehicle data cannot work given these spatiotemporal gaps. Further, Figure 4.2 gives the duration distribution of these spatiotemporal gaps. We found that 15% of the gaps have a duration at least 400s; 74% of the gaps have a duration at least 100s. The above results indicate there are a few gaps with high frequencies and long durations, which have to be addressed to improve accuracies of tracking.

4.2 HMM-based Modeling

We use one private vehicle as an example to show how coTrack tracks its locations in real time, and in practice, all private vehicles can be tracked in parallel. The input of our model is the GPS coordinates (i) generated by this private vehicle, (ii) collected and then uploaded by nearby commercial vehicles to the cloud. The output of our model is a sequence of detailed road segments to show a real-time trace of this private vehicle. A naive solution is to select the closest segment given GPS in observation and perform an interpolation when missing observations. But it has been shown by the previous work

that it fails even with small noises [37], and the interpolation for missing observations cannot work until we have a new observation.

We design our model based on HMM similar to previous work [37] [38] given its robustness to observation errors. However, our key contribution is to infer missing observations due to spatiotemporal gaps based on collaborative mobility patterns of commercial vehicles, which has not been studied before. In our model, the hidden states are locations of a private vehicle at levels of road segments with a certain length (e.g., 100 meters) given a particular time slot; the observations are a set of GPS coordinates uploaded by commercial vehicles. Figure 4.3 gives an overview of the HMM model where the space for a hidden state S_t at a time slot t are N road segments = $\{r_1, r_2, \dots, r_N\}$; an observation O_t in a slot t is a set of GPS coordinates about a vehicle.

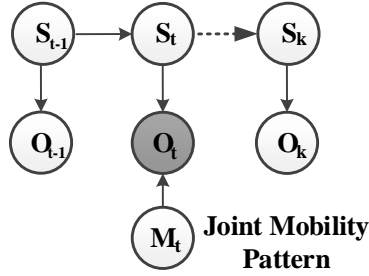


Fig 4.3: HMM Model

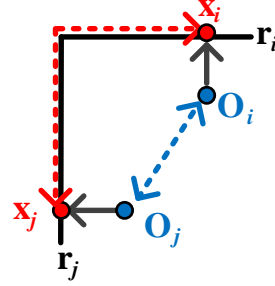


Fig 4.4: State Transition

The objective of our model is to obtain the most likely sequence of states for a series of slots given the observations. To achieve this objective, we have to decide three components: (i) the emission probability $p(O_t|S_t)$ represents the conditional probability of having this observation O_t given the vehicle being in that state; (ii) the transition probability is the probability of changing from one state (i.e., one road segment) to the next state; (iii) given these probabilities, we aim to use a traversal algorithm to find the maximum likelihood sequence of hidden states as our tracking results. We introduce them as follows.

(i) Emission Probabilities: They indicate the likelihood of an observation being made given a state S_t . An emission probability of $p(O_t|S_t = r_i)$ is decided by the distance from r_i to O_t , i.e., $Dis(r_i, O_t)$, and a Gaussian function [38].

$$p(O_t|S_t = r_i) = \frac{1}{\sqrt{2\pi}\sigma_{O_t}} e^{-0.5\left(\frac{Dis(r_i, O_t)}{\sigma_{O_t}}\right)^2}. \quad (4.1)$$

σ_{O_t} is the standard deviation of the observation O_t .

(ii) Transition Probabilities: They measure the likelihood of a vehicle traveling from a road segment candidate to another road segment candidate. Given two observations, a transition probability between two corresponding road segment candidates is higher if the travel distance between these candidate road segments is closer to the distance between two observations. For example, as in Figure 4.4, given two observations, O_i and O_j , and their respective road segment candidates r_i and r_j , let x_i be the projection of O_i on r_i , and x_j be the projection of O_j on r_j . Then, we compare the Euclidean distance between two observed O_i and O_j , i.e., $Dis(O_i, O_j)$ with the travel distance between x_i to x_j along the connected road segments, i.e., $Dis(x_i, x_j)$. The closer between two distances, the higher the transition probability. With this intuition, we quantify $p(r_i \rightarrow r_j)$ by the following.

$$(1 - \frac{|Dis(O_i, O_j) - Dis(x_i, x_j)|}{\sum_{u \in \{1, N\}} \sum_{v \in \{1, N\}} |Dis(O_i, O_j) - Dis(x_u, x_v)|}) / (N - 1). \quad (4.2)$$

The rationale is that in a short range, two projections on the correctly-matched segments have the similar distance between two observations. In practice, we do not consider the transition between two segments that cannot be traveled within a time slot at a reasonable speed, i.e., 100 m/h.

(iii) Traversal Algorithm: Given the emission and transition probabilities, we utilize the Viterbi algorithm [39], which is a dynamic programming technique to decide hidden states with the high probability. Thus we have the most likely segment sequence traveled by a private vehicle, which is used as our tracking result.

With the above three components, we successfully obtain our real-time tracking results for private vehicles. However, given the spatiotemporal gaps, the observation O_t for a particular slot for a private vehicle could be missing. Thus, as shown in Figure 4.3, our key contribution is to provide contextual information, i.e., a joint mobility pattern, to infer missing observations as given in the next subsection.

4.3 Collaborative Tracking

We design a technique based on a collaborative mobility pattern to infer missing observations in our tracking model.

4.3.1 Key Idea

The key idea of our collaborative tracking is to address spatiotemporal gaps by predicting the mobility patterns of commercial and private vehicles. Given a particular private vehicle, in every modeling time slot, we predict if this vehicle will have a gap in the next slot. This is based on its predicted mobility pattern and the predicted mobility patterns of nearby commercial vehicles. Built upon the last observation about this private vehicle, we predict lengths of the spatiotemporal gap by inferring its future encounter with any other commercial vehicle based on (i) our mobility prediction for this private vehicle and all nearby commercial vehicles including buses, taxis, and trucks; (ii) an online prediction updating (i.e., our prediction constantly changes based on real-time vehicle mobility). Based on the last observation and this predicted future encounter, we perform an interpolation to produce observations for our HMM based on road networks and real-time traffic situations. If we receive data about this private vehicle earlier or later than we predict, we adjust our prediction for this particular vehicle based on the early or later encounter time and location. Thus, we fill spatiotemporal gaps based on predicted mobility. As follows, we introduce two parts, i.e., mobility pattern prediction and online prediction updating.

4.3.2 Mobility Pattern Prediction

In this work, we first built individual mobility models for particular fleets based on their mobility features and then integrate them together for a collaborative model. Given both historical and real-time vehicle GPS data, we design a generic Bayesian model to predict next locations of vehicles by inferring their final destinations given their historical mobility patterns and real-time locations. For a vehicle k (e.g., bus, taxi, truck or a private vehicle) starting from a road segment s_k at the time t_k , we predict its

destination d_k based on historical and real-time data about this vehicle. Our model is based on an observation that human mobility patterns (in terms of origins, destinations, and starting time) are highly regular and can be learned based on historical and real-time data [27]. We rigorously examine this observation in our setting of vehicle tracking by investigating the entropy of destination $E(d_k)$ for a vehicle k and its conditional entropy $E(d_k|s_k, t_k)$ given start location s_k and time t_k for a particular trip. A lower entropy indicates a higher predictability. We calculate the entropy and conditional entropy for all fleets given one-month data based on the following.

$$\begin{aligned} E(d_k) &= - \sum_{d_k \in \Psi} p(d_k) \log p(d_k), \\ E(d_k|s_k, t_k) &= - \sum_{d_k, s_k \in \Psi, t_k \in \chi} p(s_k, d_k, t_k) \log \frac{p(s_k, t_k)}{p(s_k, d_k, t_k)}, \end{aligned} \quad (4.3)$$

where Ψ , the set of all road segments, is the support of s_k and d_k , which can be considered as random variables; χ , the set of minutes in a day, is the support of the random variable t_k . We plot the CDF of Conditional Entropy CDF in Figure 4.5.

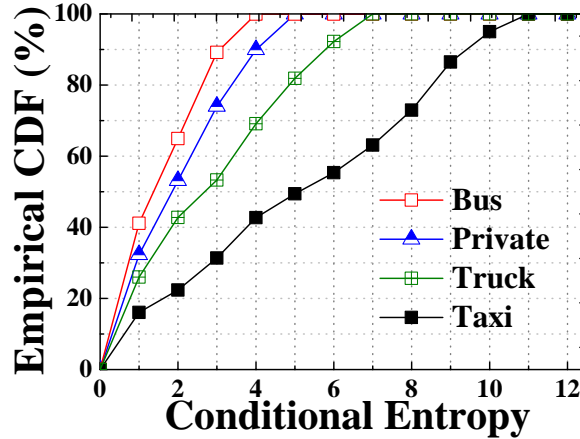


Fig 4.5: Conditional Entropy

We found that the conditional entropies of mobility patterns for buses and private vehicles are lower than trucks and taxis. For example, given the start location and start time, 53% of private trips have less than $2^2 = 4$ possible segments as destinations; 51% of taxi trips have less than $2^5 = 32$ possible segments as destinations, among all segments in Shenzhen. Based on the above analyses, we introduce our fleet-specific mobility models given their unique mobility.

Bus Pattern: Normally, the bus fleet has the most regular mobility pattern because

mobility patterns for buses are prefixed by their operating routes as shown in Figure 4.5. However, due to traffic conditions in the urban areas, buses are mostly late compared to their timetables. Thus, in this dissertation, for every bus line, we build a bus mobility model to infer next road segment in a time slot by predicting its final destination. With this model, given a specific bus, we can predict its future road segments along its route.

Truck Pattern: Compared to the bus fleet, the truck fleet has a different operating feature with a semi-regular pattern as shown in Figure 4.5. Since the truck fleet we studied is from a logistics company to deliver packages, we build a predictive model to focus on its mobility pattern. Based on discussing with the truck company, every truck basically has a region to cover, and the detailed daily route varies based on different daily requests. However, based on truck data, we found that every truck has the longest route, which uses to cover all its potential costumers' addresses. Most of its daily routes are a part of this longest route by skipping a few areas without delivery or pickup requests. As a result, we build a Bayesian model to predict a truck's future locations by inferring its next destination with its historical routes, current start time, and the current route it has passed.

Taxi Pattern: Compared to bus and truck fleets as shown in Figure 4.5, the taxi fleet has the most random pattern since their routes are decided by passengers and individual cruising patterns of drivers. However, it has been shown by the previous study, given pickup locations and time along with the route passed, an occupied taxi's future route can be predicted with a high probability given trip regularity [42]. Further, for taxis without passengers, every taxi driver has a highly regular cruising pattern given years of experiences [25]. Thus, given the trip contexts (e.g., pickup locations and times) and individual cruising patterns, we build a model to predict future routes with historical data.

Private Vehicles Pattern: It has been shown that human daily mobility pattern is highly regular [27]. Given the origin, start time, and the route passed for a trip, the final destination of this trip can be predicted with a high probability as in Figure 4.5. We use historical tracking data to build a Bayesian Model to predict routes of a private vehicle conditioning on origins, start times, and routes passed.

4.3.3 Online Prediction Updating

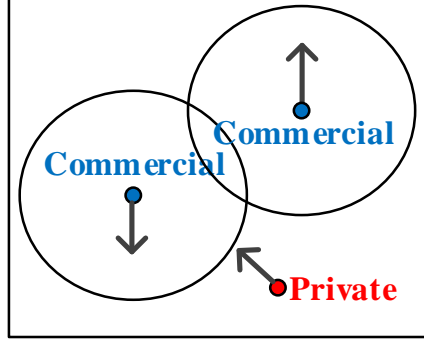


Fig 4.6: Online Updating

Given the above models for individual mobility, we integrate them into a joint mobility model to collaboratively predict future encounters of private and commercial vehicles. Our encounter prediction is an online prediction where we update our encounter prediction results based on the real-time locations of commercial vehicles. These locations can be used to implicitly infer locations of private vehicles. Figure 4.6 gives an example where a private vehicle is experiencing a spatiotemporal gap and two commercial vehicles nearby cannot track it. But since we know the constantly updated coverage areas of the commercial vehicles by their periodically GPS uploading, we can infer some road segments where this private vehicle cannot be. Otherwise, this private vehicle will be captured by the commercial vehicles. Since these coverage areas are constantly changing due to the mobility of commercial vehicles, we can update the predicted trace for this vehicle accordingly.

4.4 Summary

In this section, we provide an HMM model to infer detailed traces of private vehicles based on commercial vehicles. We design the emission and transition probabilities and utilize the classic Viterbi algorithm to decide the hidden states with the high probability to obtain tracking results. Finally, we address a key challenge of missing observations by developing mobility models for these vehicular fleets based on their individual mobility patterns.

Chapter 5

Mobility Modeling Layer: Field Study

In this section, we introduce our preliminary deployment for a road test to evaluate the performance of the current cellular and onboard infrastructures with a 3G service plan in terms of data uploading for real-time tracking.

Although most of the commercial fleets in Shenzhen have communication devices already for periodical data uploading, they are focused on centralized communications mostly for accounting, instead of real-time tracking. Therefore, only small amount of data about a vehicle itself, e.g., its GPS records, is uploaded periodically with a 3G plan for a small monthly fee. But in our coTrack, a vehicle has to upload data of both itself and nearby vehicles. Thus, it is unclear if current uploading infrastructures in urban mobile environments can support our real-time tracking or not, without upgrading to a more expensive 4G service.

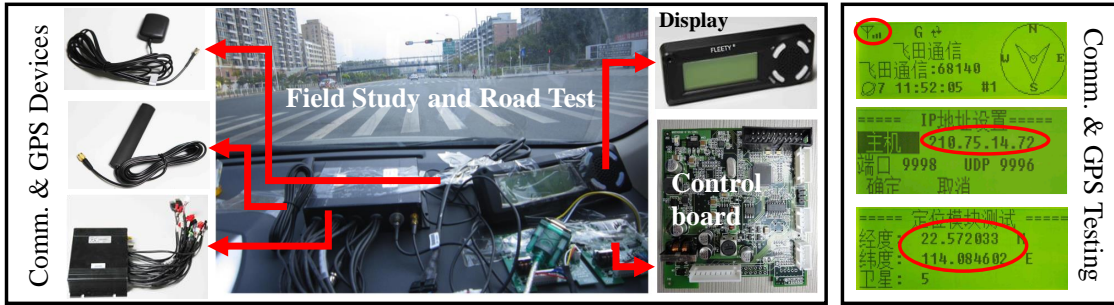


Fig 5.1: Initial Deployment and Road Test for Tracking Record Uploading

Based on our collaboration with Shenzhen Transportation Committee, we instrumented a small portion (106 taxis) of the taxi fleet with 14 thousand taxis in Shenzhen for our vehicular sensing project. The instrumentation includes a GPS module, a communication module for both peer-to-peer and centralized communications, a central control system (STM32F103), and a display. A preliminarily-instrumented taxi for

our road test is given in Figure 5.1, along with GPS and communication configuration screenshots. We use this road test to evaluate real-time data uploading speeds for our centralized vehicle tracking. During the road test, different sizes of status records for a commercial vehicle were generated to simulate different densities of private vehicles. We repeatedly uploaded 10 tracking records with different sizes and the average uploading speeds are given in Figure 5.2.

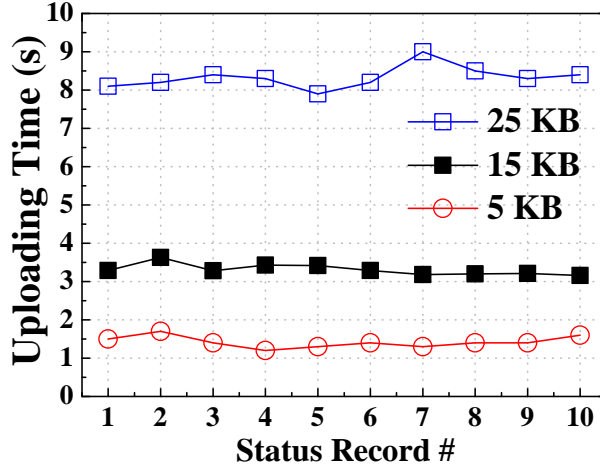


Fig 5.2: Mobile Tracking Record Uploading

The results show that for a 5KB record, it takes less than 2 seconds to upload, which can contain information for more than 100 private vehicles in its communication range. For a denser scenario where 300 vehicles in a 100-meter radius, we can receive a 15KB record within 4 seconds. Even for an extremely dense scenario, i.e., 500 vehicles in a 100-meter radius, a record of 25KB can be uploaded within 9 seconds. These results validate the commercial fleets and cellular networks in Shenzhen are ready for real-world tracking data uploading in the current data plans.

As for an urban-scale tracking with thousands of vehicles based on peer-to-peer communication, we currently do not have a dense network of instrumented vehicles for a field test. The main reason is that we are having some difficulties to instrument private vehicles with peer-to-peer communication (mainly due to lacking incentives) without the mandatory order from the government. Instead, in the next section, we perform a trace-driven evaluation for coTrack.

Chapter 6

Mobility Modeling Layer: Evaluation

We introduce our trace-driven evaluation for coTrack in terms of data management, methodology and results.

6.1 Evaluation Data Management

Since we concentrate on the evaluation on the coTrack system, we briefly introduce issues related to fleet data management for analyses and evaluations given the space limitation. Based on a secure and reliable transmission mechanism, our server is fed with data from STC, truck and navigation companies with a wired connection. We have been storing a large amount of data to investigate these urban fleets as in Figure 3.1. Such a large amount of streaming mobility data requires significant efforts for the daily management. For data processing, we utilize a high performance cluster with both Hadoop and Spark platforms including (i) 10 Dell machines with 4 Tesla K80c each; (ii) 12 HP machines with 2 Tesla K80c each; (iii) 4 Xeon E5-2650 with a half TB memory each; (iv) A series of 800GB SSD and 15TB of spinning-disk spaces; (v) 2 PB additional disk space. Due to the extremely large size of our data and their streaming nature, we have been finding duplicated data, missing data and data with logical errors. So we have been conducting a detailed cleaning process to filter out errant data on a daily basis. We protect the privacy of drivers by anonymizing all data and details are given in § 8.

6.2 Evaluation Methodology

Based on one month of data introduced in Figure 3.1, we introduce key components related to our evaluation including ground truths, evaluation metrics, baseline approaches,

and impacts of real-world and system parameters.

Ground Truths: In this dissertation, we obtain ground truths of private vehicle trajectories with their uploaded GPS data obtained by onboard navigators as introduced in Section 3. This set of 10 thousand private vehicles (293 thousand nationwide) accounts for 0.5% of all 2 million vehicles in Shenzhen. As shown in Figure 3.2, this private vehicle fleet covers major road segments in Shenzhen compared to other fleets, and can be used as a representative set of all private vehicles in Shenzhen. Based on our discussion with onboard navigator users and the company providing data, we confirm that even for daily commutes where drivers are familiar with routes, they are still likely to turn on navigators. This is because it has real-time traffic conditions and it can detect locations of traffic violation cameras. If the navigators are turned on, GPS locations of vehicles are uploaded for traffic speed monitoring and potential location-based services.

Evaluation Metrics: In this dissertation, we utilize similarity between tracking results and the ground truth of vehicle locations to test the performance of our coTrack system. We quantify the similarity between two trajectories (one is our result, and the other is the ground truth) by calculating distances between them [22]. Given two trajectories:

$$\begin{aligned} t_1 &= [p_{1,1}, p_{1,2}, \dots, p_{1,n}]. \\ t_2 &= [p_{2,1}, p_{2,2}, \dots, p_{2,n}]. \end{aligned} \tag{6.1}$$

To measure the distance between t_1 and t_2 , we consider each point $p_{1,i}$ in t_1 and calculate its distance from its corresponding point in t_2 . We find the corresponding point by obtaining the point with the closest time of $p_{1,i}$ in t_2 . Note that the time of $p_{j,i}$ is measured by the average time interval obtained by the total travel time and the number of points. Then we calculate the Euclidean distance between these two points, i.e., $p_{1,i}$ and its corresponding point in t_2 .

$$dist(t_1, t_2) = avg(d(p_{1,i}), d(p_{2,i})), \forall p_{1,i} \in t_1, p_{2,i} \in t_2. \tag{6.2}$$

If two trajectories have different time intervals, we exchange t_1 and t_2 and then calculate $dist(t_2, t_1)$. Finally, the distance is given by the following

$$dist = max(dist(t_1, t_2), dist(t_2, t_1)). \tag{6.3}$$

When we evaluate the results of coTrack in each hour, we calculate the average distance between our results and the ground truth for all private vehicles in that hour. Finally, we normalize this distance to obtain our metric Normalized Average Distance Error (NADE).

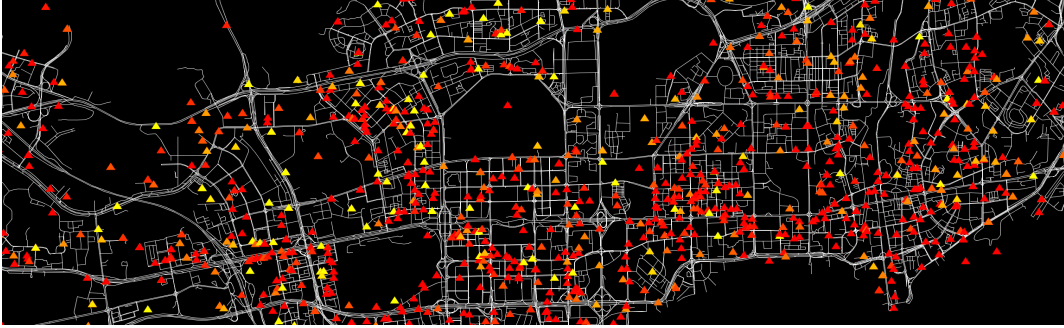


Fig 6.1: Private Vehicle Density on Cell Tower Levels

Baseline Approaches: As follows, we introduce two baseline approaches for performance comparisons.

(i) **STrack:** We compared coTrack with an approach based on static urban infrastructures to track private vehicles in major intersections. This fixed baseline system STrack represents a wide range of infrastructure-based systems, e.g., cameras or RFID, to track urban-scale private vehicles. In STrack, we envision fixed devices, e.g., camera, RFID, or ride-side unit for DSRC devices, are deployed in major intersections of Shenzhen road networks to track private vehicles. We implement STrack by assuming that a given percentage (50% in default) of intersections have been installed a device that can track private vehicles when they are passing intersections. We also evaluate the impact of percentages of intersections with infrastructures on STrack. In contrast, our coTrack system is based on a mobile infrastructure approach and have mobile coverage.

(ii) **CTrack:** It is a state-of-the-art system [37] to track individual vehicles based on cellular networks by periodical communications between onboard cellphones and cell towers, and then to use the locations of observed cell towers to infer locations of cellphones and thus vehicles. We implement CTrack based on cell tower locations of a cellular network in Shenzhen by assuming every private vehicle has a cellphone to track a vehicle’s location on cell tower levels. Figure 6.1 gives a visualization of private

vehicles density on cell tower levels where we assign private vehicles to the closest cell towers based on their GPS locations. To make CTrack more competitive, we use private vehicle GPS to obtain locations and time of private vehicles making turns, which is to simulate an optional function of CTrack where various smartphone sensors are used to infer the turns of the vehicles to increase tracking accuracies.

Impacts of Factors: We evaluate three real-world factors and their impacts on the performance of coTrack. **(i) Fleet Types:** To investigate the impact of different fleets on our systems, we feed coTrack with four different commercial fleets, i.e., bus, taxi, logistic truck, and the combination of them. **(ii) Locations and Day of Week:** We evaluate the performance of coTrack at different regions of the city, i.e., downtown and suburban districts, and the day of the week, i.e., weekend and weekday, where the mobility patterns of both commercial and private vehicles are different. **(iii) Tracking Parameters:** Finally, three parameters have significant impacts on vehicle tracking based on peer-to-peer communications: the tracking range, the tracking duration, and the tracking probability, i.e., the probability of tracking a private vehicle by a commercial vehicle if they are in the tracking range longer than the tracking duration. The default settings for them are 200m, 10s, 100% and 8 AM.

6.3 Evaluation Results

We first compare our coTrack system with two baseline approaches, and then investigate the impacts of three sets of real-world factors on coTrack performance, i.e., fleet types, locations and time of the week, and tracking parameters, and finally we present a detailed summary of our evaluation.

6.3.1 Baseline Comparison

We evaluate coTrack by comparing it to STrack and CTrack, and Figure 6.2 gives the results of these systems.

Figure 6.2, the X axis is the time of day, and Y axis is our metric NADE for every hour. Based on the results, we found that coTrack has better performance than STrack

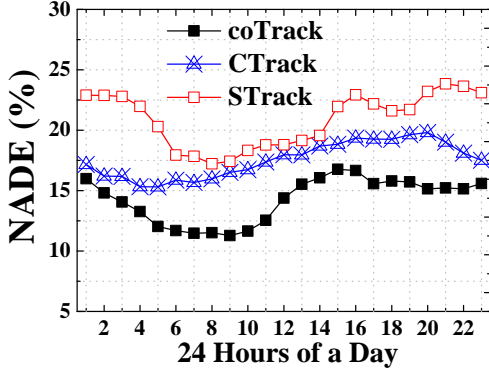


Fig 6.2: Baselines

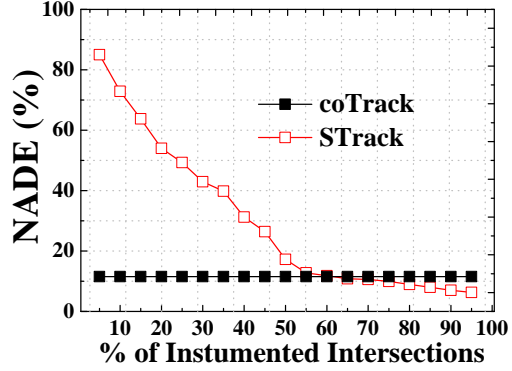


Fig 6.3: % of Infrastructures

during 24 hours of a day with an average performance gain of 42.2%. This is due to the fact that coTrack utilizes a virtual mobile infrastructure, which has a more flexible coverage. Moreover, coTrack outperforms CTrack by 23.2% on average because of its mobile nature, and CTrack only utilized fixed cell towers to track private vehicles, which limits its tracking range. Finally, coTrack utilizes the existing infrastructure in the commercial vehicles, while implementing both STrack and CTrack needs significant new infrastructure investment, e.g., asking every driver to install an app to detect nearby cell towers or deploying a large number of cameras or road side units. To show the impacts of STrack infrastructure scales on the results, we vary the percentage of intersections with tracking devices and results are given in Figure 6.3. We found that with the increase of tracking devices in intersections, the performance of STrack becomes better. But even with a 60% deployment rate, STrack still cannot outperform coTrack due to its static nature.

6.3.2 Impacts of Fleet Types

We first apply coTrack with 4 different fleet types, which are bus, taxi, logistics truck, and the combination of them. We call them as coTrack-B, coTrack-T, coTrack-L and coTrack respectively. Figure 6.4 plots the NADE of these 4 versions of coTrack. In general, the accuracy of coTrack is better than the other three models. This is because the performance of coTrack is dependent on the encounter frequency between private vehicles and commercial vehicles. In particular, comparing three individual fleet-driven

versions, i.e., coTrack-T, coTrack-B and coTrack-L, the accuracy of coTrack-T is better than coTrack-B and coTrack-L. This is because the mobility pattern of the taxi fleet is more diverse compared to the bus fleet and logistics truck fleet, which leads to a larger coverage. It can also be validated by our fleet visualizations and entropy analyses in Figures 3.2 and 4.5. In addition, during the daytime, coTrack is more reliable since there are more commercial vehicles operating, compared to other individual fleets.

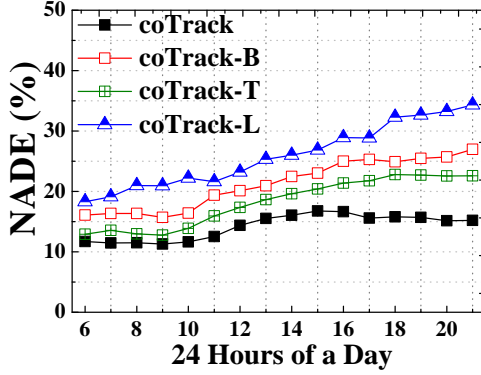


Fig 6.4: Impacts of Fleets

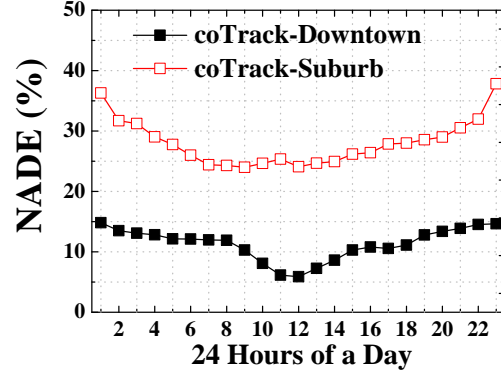


Fig 6.5: Impacts of Location

6.3.3 Impacts of Locations and Time

In this subsection, we investigate the performance of coTrack on two urban districts, i.e., a downtown district and a suburb district, among eleven districts in Shenzhen to investigate the impacts of locations. These two districts also indicate different combinations of geographic and demographic features. Figure 6.5 shows the performance of coTrack in the downtown and suburb districts. The X axis is time of day, and Y axis is our metric NADE. We found that in general coTrack performs better in the downtown district compared to its performances in the suburb district. This is because for coTrack the downtown district has more commercial vehicles to track private vehicles compared to the suburb district. As a result, more commercial vehicles lead to better spatiotemporal coverage, thus better tracking performance. Further, to investigate the impacts of day of the week, we plot the performances of coTrack on the weekday and weekend in Figure 6.6.

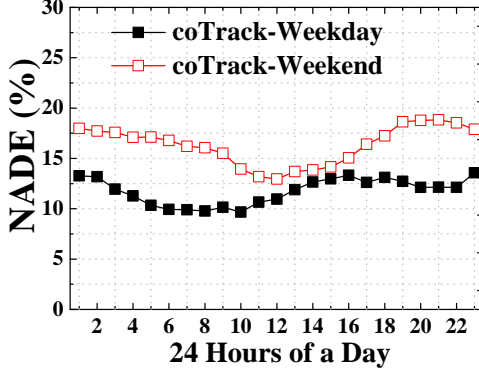


Fig 6.6: Impacts of Week

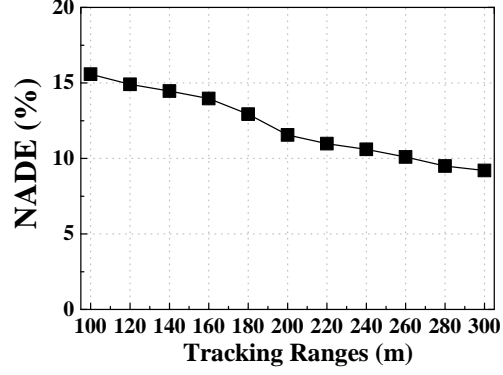


Fig 6.7: Impact of Ranges

We found that in general coTrack-Weekday has better performance than coTrack-Weekend. This is because during the weekday, we have more commercial vehicles to track private vehicles mainly used for daily commutes. On the weekend, coTrack has the best performance from 10 AM to 12 PM; whereas on the weekday, the best performance is around 6 AM -9AM, correlating to general traffic patterns.

6.3.4 Impact of Tracking Parameters

In this section, we investigate the performance of our system with respect to important tracking parameters, i.e., the tracking range, durations and probabilities.

Impacts of Ranges: We change the tracking range from 100 meters to 300 meters to see the performance of coTrack. Figure 6.7 plots the NADE of coTrack under different tracking ranges. We found that in general, the longer the tracking ranges, the better the performance, e.g., coTrack has the best performance with the tracking range of 300 meters. The reason for this improvement is that when the tracking range is longer, there are more private vehicles that can be tracked by commercial vehicles given the same mobility patterns.

Impacts of Durations: The tracking duration indicates the time period during which a commercial vehicle can track a private vehicle if the distance between them is shorter than the predefined tracking range. If they are in the tracking range shorter than the predefined tracking durations, the commercial vehicle cannot track the private vehicle in our setting. We change the tracking duration from 2 seconds to 30 seconds, and

then plot the results in Figure 6.8. We found that in general, the shorter the tracking duration, the better the performance, e.g., coTrack has the best performance with the tracking duration of 2 seconds. However, in the real world, this parameter is varied based on many factors, e.g., interferences, radio types, demodulation, etc.

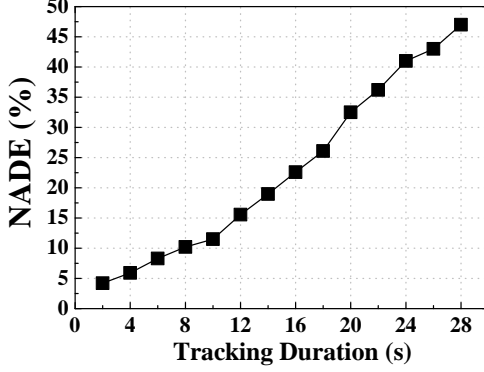


Fig 6.8: Impacts of Duration

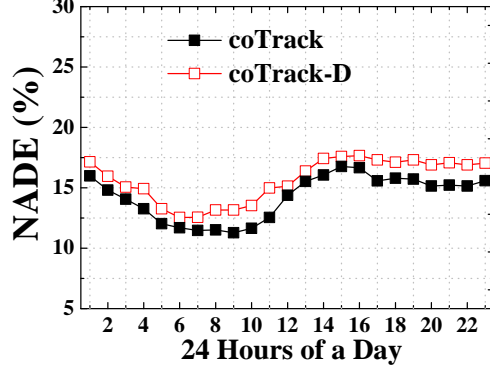


Fig 6.9: Impacts of Prob.

Impacts of Probability: We change the tracking probability according to the distance of two vehicles by a Gaussian function. Figure 6.9 shows the comparison between coTrack with fixed tracking probability value 1 and a version of coTrack with a dynamic probability, called coTrack-D. We found that although in coTrack-D, a vehicle has a lower chance to be tracked by a commercial vehicle due to a dynamic probability based on their distance, the performances of coTrack and coTrack-D are similar. A possible explanation is that for the most of the time, we have enough commercial vehicles to track private vehicle redundantly. So a dynamic probability does not have a major impact on the performance of our model.

6.4 Evaluation Summary

We have following observations. (i) As shown in Figures 6.2 and 6.3, coTrack has a better performance in general than two systems based on fixed infrastructures at major intersections and cellphone networks, which indicates a mobile tracking system has better performance than a fixed tracking system. (ii) As shown in Figure 6.4, different types of commercial vehicles have significant impacts on coTrack. In general, using more commercial fleets leads to a lower NADE due to increased numbers of commercial

vehicles for tracking. (iii) As shown in Figures 6.5 and 6.6, different urban locations and day of the week have different impacts on the performance of coTrack. The tracking performance in the downtown area during the weekday is better than the performance in the suburb area during the weekend due to the diversity of fleets and commuting patterns. (iv) As shown in Figures 6.7, 6.8 and 6.9, for three tracking parameters, it shows tracking durations and ranges have higher impacts than distance-based tracking probabilities on the performance of coTrack given the high density of commercial fleets.

Chapter 7

Application Layer

We motivate, design, and evaluate a real-world service where we utilize private vehicle tracking to infer travel time.

7.1 Application Background

Inferring urban-scale travel time in a real-time fashion is essential to navigation services, e.g., Google Maps [11] and Apple Maps [2]. Currently, these map services mainly use large-scale historical data and small-scale real-time data from a limited number of vehicles to infer travel time given origins and destinations. This is because they do not have enough sampling data to infer real-time traffic speeds at urban scale [38]. In this section, we focus on using our private vehicle tracking results for travel time between different urban regions for interregional mobility detections. Inferring inter-regional mobility is an important aspect of urban transportation since it reflects traffic congestion and human mobility on region levels. The inter-regional travel time typically varies based on real-world contextual information, e.g., time of day, day of week, social events, accidents, etc. To detect such travel time in real time, the existing work is based on devices and infrastructures including smartphones [37] or call detail records [29]. However, these approaches may not work in real-time at urban scale because of the limited penetration rates. Some services have also been proposed to estimate travel time based on commercial vehicles, e.g., taxis [16]. But due to the limited number of taxis, real-time travel time between large-scale of regions is still inferred based on historical data. In contrast, we propose a service to detect mobility in terms of travel time using private vehicles, which have better coverage than commercial vehicles due to their large volumes.

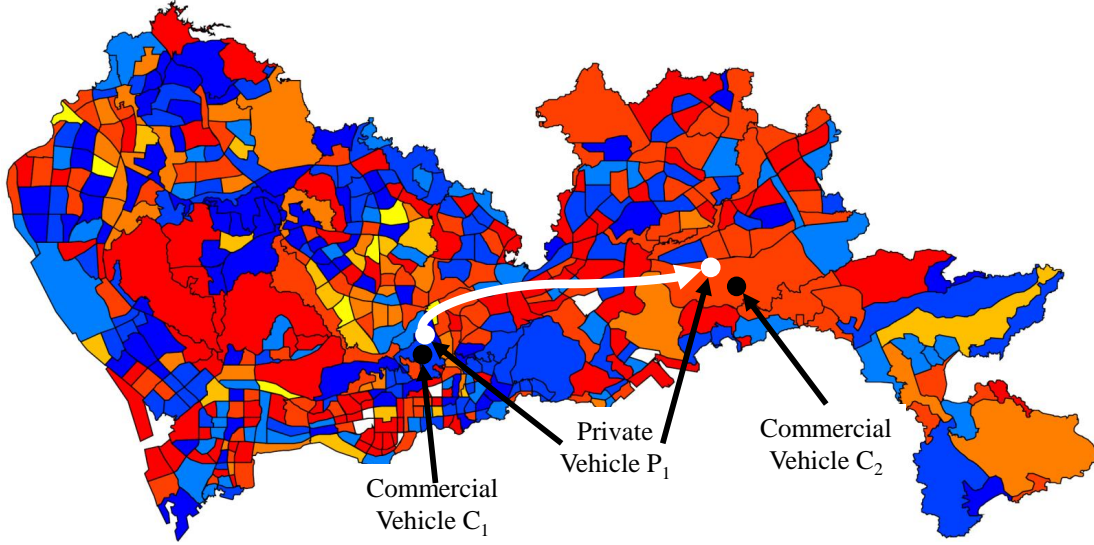


Fig 7.1: Interregional Mobility

For example, in Figure 7.1, if (i) a private vehicle P_1 is detected by a commercial vehicle C_1 at one region at time t_1 , and (ii) the same private vehicle P_1 is tracked by another commercial vehicle C_2 at another region at time t_2 , then we can use $t_2 - t_1$ to infer the real-time travel time between these two regions even if we do not have other commercial vehicles tracking P_1 from t_1 to t_2 . This result can then be improved based on a large number of private vehicles traveling between these two regions and tracked by commercial vehicles. Also, it has been shown that the drivers of commercial vehicles have different driving patterns from the drivers of private vehicles, e.g., driving speeds, route selections, and lane changes [41]. Thus, our travel time inferred by private vehicles is more useful for other drivers of private vehicles than services based on commercial vehicles.

7.2 Application Evaluation

To evaluate our service, we first divide Shenzhen into 491 urban regions according to an administrative partition of Shenzhen. Figure 7.1 gives a visualization of the 491 regions where the blue color means there are more incoming vehicles and the red color means there are more outgoing vehicles during the morning rush hour 8 AM to 9 AM of a weekday. The deeper the color, the more the vehicles.

Different from grid-based partitions used in many dissertations [32] [16], our partition considers both geographical and demographical features. In k time slots, e.g. $\frac{24 \times 60}{5}$ five-minute slots of a day, there are $491 \times 491 \times k$ sets of travel time that need to be detected based on private vehicles. From our mobility modeling layer, we use our tracking results to map private vehicles into 491 regions and then calculate average travel time between different regions for different slots. At each time slot, we compute how long it takes for a vehicle to travel from one region to another region. Since different vehicles may start from different locations and select different paths, we choose the mean value to represent the expected travel time from one region to another region. Finally, we group the results based on travel start time. Our objective is to obtain the real-time travel time estimations between region pairs as many as possible. The more region pairs we have, the higher coverage that we can obtain at urban scale. We compare our service to a taxi-based approach [16] as a baseline.

We utilize a spatial performance gain $G = \frac{RP_{PrivateOnly}}{RP_{Total}}$ to quantify the performance of our approach where $RP_{PrivateOnly}$ is the number of region pairs detected by private vehicles only; where RP_{Total} is the number of region pairs detected by both private vehicles and taxis. Figure 7.2 gives the distribution of G by 24 hours in five-minute slots.

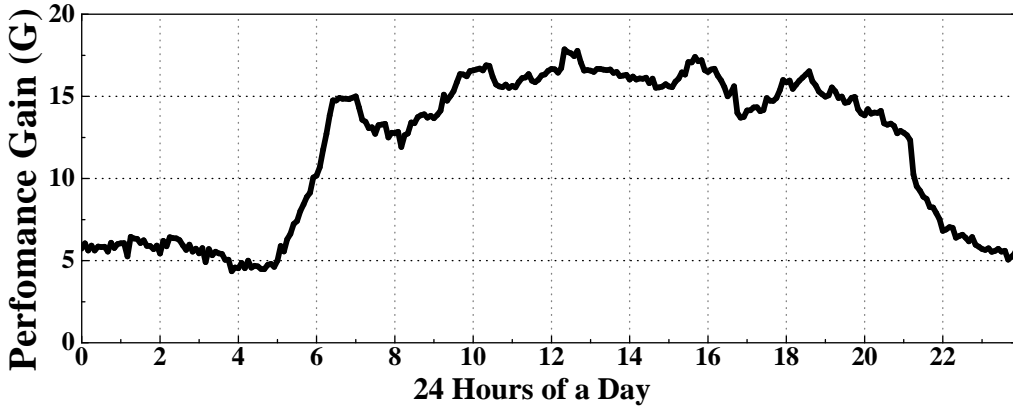


Fig 7.2: Performance Gain

We found that with diversity patterns of private vehicles we can infer more than 15% of region pairs during the regular daytime from 7AM to 8PM where more private vehicles are traveling between regions providing diverse travel patterns. Even during the night

time or early morning, e.g., from 8PM to 6AM, we can still infer 5% more region pairs when using private vehicle traces. This is because Shenzhen is a large city of 2,000 km², and some suburban regions are often without any taxis during certain hours, and our private vehicle tracking can help to provide travel time estimations starting or ending within these regions.

Given the limited number of private vehicles we studied, i.e., 0.5%, it suggests the performance gain would be higher if we have more vehicles involved. But we cannot verify it based on the current scale of our private vehicles.

Chapter 8

Discussion

Limitations. Even with a trend of the government mandating [9] [4], a major limitation of our coTrack system is to require private vehicles to broadcast its status. However, we argue that the design philosophy of coTrack, i.e., utilizing a small number of well-equipped vehicles to track a large number of minimally-equipped vehicles at urban-scale in real time, can be generalized to other scenarios. For example, a commercial vehicle can use cameras to track the vehicles nearby and upload these images to a cloud server for centralized tracking. In this case, our system design can also be applied, and the only component requiring major changes is the infrastructure layer in Figure 2.1.

Privacy Protections. While vehicle tracking has the potential for great social benefits, we have to protect the privacy of drivers involved. We took the following active steps for privacy protections. (i) De-identification : All data analyzed are anonymized by the staff of service providers, and all identifiable IDs are replaced by a serial identifier during the analyses. (ii) Minimal Exposure: We only store and process data that are useful for our vehicle tracking project, and drop other information for the minimal exposure. (iii) Aggregation: the tracking results obtained by coTrack are given at aggregated results at segment levels in a time duration instead of street addresses for a specific timestamp.

Public Data Access. Accessing empirical fleet datasets is vital to the vehicle tracking research, but such datasets are usually not available for the fellow researchers due to the various real-world issues. As an initial step, our collaborators have agreed that the data used in this work can be made for public access with some proper preprocessing. Thus, we will release our fleet data with privacy protection schemes after we can reveal

our identity.

Implementation in Different Cities. The private vehicles in different cities typically have different mobility patterns due to geographic and demographic features. It is therefore extremely important to implement coTrack in different cities. To validate its cross-city performances, we have access to a small fleet of vehicles in Beijing, including private vehicles, taxis and trucks and we are negotiating with other service providers for more data to implement coTrack.

Chapter 9

Conclusion

In this dissertation, we have identified the practical challenges in combining centralized sensing and distributed sensing together for real time vehicular sensing. We have presented viable approaches to address these challenges and demonstrated the feasibility of using a small number of well-equipped commercial vehicles to sense a large number of minimally-equipped private vehicles in real time through our hybrid sensing framework.

The main contribution of this dissertation is to explore the design of the combination of centralized sensing and distributed sensing and provide a hybrid vehicular sensing framework to enable real time vehicular sensing in an urban scale by a small number of commercial fleets.

Our endeavors offer a few valuable insights:

- the commercial vehicles provide high spatiotemporal urban coverage due to their operating features, leading to high potential for urban phenomenon modeling;
- under extremely fine-grained partitions, the commercial vehicles experience spatiotemporal gaps, which lead to inaccuracies for tracking;
- given the complementary mobility patterns, the heterogenous commercial vehicles can address these spatiotemporal gaps with a collaborative method to track private vehicles;
- our work only addresses the technical frontier on the modeling part, and it is even more critical to establish a right policy that would make large-scale deployment feasible.

These insights consolidate our thesis that:

- the high spatiotemporal urban coverage enables the commercial fleets to collect enough data from private vehicles in real time;
- even though there exist spatiotemporal gaps, they still can be addressed by the sufficient sensing data in the previous minutes;
- with the study of previous collected and analyzed mobility pattern of vehicles, these spatiotemporal gaps are predictable;
- the high spatiotemporal urban coverage and these solvable and predictable spatiotemporal gaps make sure our hybrid sensing framework to be feasible.

Through our hybrid sensing framework, we utilize those underutilized resource equipped in vehicles without enlarging the burden of a city. With the prediction of spatiotemporal gaps, we can rearrange the daily routes of commercial vehicular fleets to fill up them and make some services become viable, such as collecting video data from vehicular camera such kind of data with big size in real time and sharing medium information from commercial fleets to private vehicles.

References

- [1] 9 delightfully geeky stats about nyc bridges and tunnels. <https://www.buzzfeed.com/johntemplon/9-delightfully-geeky-stats-nyc-bridges-tunnels>.
- [2] Apple map. <http://www.apple.com/ios/maps>.
- [3] Bmw assist. <http://www.bmwusa.com/standard/content/explore/bmw-value/bmwassist/default.aspx>.
- [4] Cadillac to soon roll out v2v safety tech in cts in canada. <http://canada.autonews.com/article/20170309/canada01/303099941/cadillac-to-soon-roll-out-v2v-safety-tech-in-cts-in-canada-u-s>.
- [5] Connected ann arbor. <https://www.its.dot.gov/factsheets/dsrcfactsheet.htm>.
- [6] Dot proposes mandating cars broadcast location, direction and speed. <http://www.cnsnews.com/commentary/terence-p-jeffrey/dot-proposes-mandating-cars-broadcast-location-direction-and-speed>.
- [7] Dsrc: The future of safer driving. <https://www.its.dot.gov/factsheets/dsrcfactsheet.htm>.
- [8] Ezpass. <https://www.e-zpassny.com/en/about/plazas.shtml>.
- [9] Feds want v2v communication in new cars starting in 2021. <http://blog.caranddriver.com/feds-want-v2v-communication-in-new-cars-starting-in-2021>.
- [10] Garmin navigator. <https://explore.garmin.com/en-us/nuvi>.
- [11] Google map. <https://www.google.com/maps>.
- [12] New york city camera. <http://dotsignals.org>.
- [13] Two billion vehicles projected to be on roads by 2035. <http://www.csmonitor.com/business/in-gear/2014/0729/two-billion-vehicles-projected-to-be-on-roads-by-2035>.
- [14] Want to drive in beijing? good luck in the license plate lottery. <https://www.nytimes.com/2016/07/29/world/asia/china-beijing-traffic-pollution.html>.
- [15] ASLAM, J., LIM, S., PAN, X., AND RUS, D. City-scale traffic estimation from a roving sensor network. In *Proceedings of 10th ACM Conference on Embedded Network Sensor Systems, SenSys '12*.

- [16] BALAN, R. K., NGUYEN, K. X., AND JIANG, L. Real-time trip information service for a large taxi fleet. In *Proceedings of the international conference on Mobile systems, applications, and services*, MobiSys '11.
- [17] BIAGIONI, J., AND ERIKSSON, J. Map inference in the face of noise and disparity. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '12.
- [18] BIAGIONI, J., GERLICH, T., MERRIFIELD, T., AND ERIKSSON, J. Easytracker: Automatic transit tracking, mapping, and arrival time prediction using smart-phones. SenSys '11.
- [19] CAPURSO, N., ELSKEN, E., PAYNE, D., AND MA, L. Poster: A robust vehicular accident detection system using inexpensive portable devices. In *MobiSys* (2014).
- [20] CHEN, M. Y., SOHN, T., CHMELEV, D., HÄHNEL, D., HIGHTOWER, J., HUGHES, J., LAMARCA, A., POTTER, F., SMITH, I. E., AND VARSHAVSKY, A. Practical metropolitan-scale positioning for gsm phones. In *UbiComp* (2006).
- [21] CHENG, B., ROSTAMI, A., AND GRUTESER, M. Experience: Accurate simulation of dense scenarios with hundreds of vehicular transmitters. In *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking* (New York, NY, USA, 2016), MobiCom '16, ACM, pp. 271–279.
- [22] CUDRE-MAUROUX, P., WU, E., AND MADDEN, S. Trajstore: An adaptive storage system for very large trajectory data sets. In *Data Engineering (ICDE), 2010 IEEE 26th International Conference on* (2010), IEEE, pp. 109–120.
- [23] GANTI, R. K., PHAM, N., AHMADI, H., NANGIA, S., AND ABDELZAHER, T. F. Greengps: A participatory sensing fuel-efficient maps application. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services* (New York, NY, USA, 2010), MobiSys '10, ACM, pp. 151–164.
- [24] GE, Y., LIU, C., XIONG, H., AND CHEN, J. A taxi business intelligence system. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11.
- [25] GE, Y., XIONG, H., TUZHILIN, A., XIAO, K., GRUTESER, M., AND PAZZANI, M. An energy-efficient mobile recommender system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (2010), KDD '10.
- [26] GONZALEZ, H., HAN, J., LI, X., MYSLINSKA, M., AND SONDAK, J. P. Adaptive fastest path computation on a road network: a traffic mining approach. In *Proceedings of the 33rd international conference on Very large data bases* (2007), VLDB '07.
- [27] GONZALEZ, M. C., HIDALGO, C. A., AND BARABASI, A.-L. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779–782.
- [28] HO, B.-J., MARTIN, P. D., SWAMINATHAN, P., AND SRIVASTAVA, M. B. From pressure to path: Barometer-based vehicle tracking. In *BuildSys@SenSys* (2015).

- [29] ISAACMAN, S., BECKER, R. A., CÁCERES, R., MARTONOSI, M., ROWLAND, J., VARSHAVSKY, A., AND WILLINGER, W. Human mobility modeling at metropolitan scales. In *MobiSys* (2012).
- [30] KJÆRGAARD, M. B., JENSEN, J. L., GODSK, T., AND TOFTKJÆR, T. Entracked: energy-efficient robust position tracking for mobile devices. In *MobiSys* (2009).
- [31] LIU, W., LIU, J., JIANG, H., XU, B., LIN, H., JIANG, G., AND XING, J. Wilocator: Wifi-sensing based real-time bus tracking and arrival time prediction in urban environments. In *2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)* (2016).
- [32] MIAO, F., LIN, S., MUNIR, S., STANKOVIC, J., HUANG, H., ZHANG, D., HE, T., AND PAPPAS, G. J. Taxi dispatch with real-time data in metropolitan areas. *ACM ICCPS* 2015.
- [33] NANDUGUDI, A., KI, T., NUESSELE, C., AND CHALLEN, G. Pocketparker: Pocketsourcing parking lot availability. *UBICOMP '14*.
- [34] SEN, R., AND BALAN, R. K. Challenges and opportunities in taxi fleet anomaly detection. *SENSEMININE'13*.
- [35] SMIRNOV, A. V., KASHEVNIK, A., PONOMAREV, A., AND SHILOV, N. Location-based on-board system for e-tourism: Implementation for ford sync. In *ICCVE* (2014).
- [36] THIAGARAJAN, A., BIAGIONI, J., GERLICH, T., AND ERIKSSON, J. Cooperative transit tracking using smart-phones. In *SenSys* (2010).
- [37] THIAGARAJAN, A., RAVINDRANATH, L., BALAKRISHNAN, H., MADDEN, S., AND GIROD, L. Accurate, low-energy trajectory mapping for mobile devices. In *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation* (Berkeley, CA, USA, 2011), NSDI'11, USENIX Association, pp. 267–280.
- [38] THIAGARAJAN, A., RAVINDRANATH, L., LACURTS, K., MADDEN, S., BALAKRISHNAN, H., TOLEDO, S., AND ERIKSSON, J. Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones. In *SenSys* (2009).
- [39] VITERBI, A. J. A personal history of the viterbi algorithm. In *IEEE Signal Processing Magazine* (2006).
- [40] YUAN, J., ZHENG, Y., AND XIE, X. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (2012), KDD '12.
- [41] YUAN, J., ZHENG, Y., XIE, X., AND SUN, G. Driving with knowledge from the physical world. In *Proceedings of the international conference on Knowledge discovery and data mining*, KDD '11.
- [42] YUAN, J., ZHENG, Y., ZHANG, C., XIE, W., XIE, X., SUN, G., AND HUANG, Y. T-drive: driving directions based on taxi trajectories. In *GIS* (2010).

- [43] ZHAO, M., YE, T., GAO, R., YE, F., WANG, Y., AND LUO, G. Vetrack: Real time vehicle tracking in uninstrumented indoor environments. In *SenSys* (2015).
- [44] ZHENG, Y., CAPRA, L., WOLFSON, O., AND YANG, H. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 3 (2014), 38.
- [45] ZHENG, Y., CHEN, Y., LI, Q., XIE, X., AND MA, W.-Y. Understanding transportation modes based on gps data for web applications. *ACM Trans. Web* 4, 1 (Jan. 2010).
- [46] ZHOU, P., CHEN, Z., AND LI, M. Smart traffic monitoring with participatory sensing. In *SenSys* (2013).
- [47] ZHOU, P., ZHENG, Y., AND LI, M. How long to wait? predicting bus arrival time with mobile phone based participatory sensing. *IEEE Trans. Mob. Comput.* 13 (2012), 1228–1241.