Modeling Urban Trip Demands in Cloud-Commuting System: A Holistic Approach

Guanxiong Liu #, Menghai Pan #, Yanhua Li #, Zhi-Li Zhang †, Jun Luo ‡

Abstract-Rapid pace of global urbanization has posed significant challenges to urban transportation infrastructures. Existing urban transit systems suffer many well-known shortcomings, where public transits have limits on coverage areas, and fixed schedules, and private transits are expensive and fail to timely meet the demand needs. We thus envision a Cloud-Commuting system, that employs a giant pool of centralized taxis/shuttles to better cope with the dynamic urban trip demands. To better understand the feasibility of such a system, in this paper we develop generative models to capture fundamental demand arrival and service patterns, and introduce a novel model to estimate the total number of vehicles needed to serve all urban demands. We conduct experiments using large scale urban taxi trajectory data from Shenzhen, China, and compare our proposed models with empirical baselines. We obtained promising results, which shed great lights on future smart transportation system designs.

Index Terms—Cloud-Commuting, urban computing, queuing theory

I. INTRODUCTION

With the fast pace of global urbanization, the world urban population has reached 54% in 2014, and it is projected that by 2050, two-thirds of the world population will be urban [12]. The rapid growth in urban population has placed an enormous strain on urban transportation infrastructures, and has worsen the urban traffic congestion problem significantly.

Nowadays, there are two primary modes of urban transit services, i.e., public transit services such as buses, subway, and private transit services such as taxis, shared-van shuttles, ridehailing services (e.g., Uber or Lyft). Public transits run along fixed routes with fixed time tables, and have limited coverage areas, while the private transits are expensive and largely "ondemand", thus the supplies may not meet the demands in a timely manner. To tackle these challenges of the current urban transit systems, we envisage a forward-looking Cloud-Commuting based transit system for smart cities. Each vehicle is controlled by centralized servers. When a trip demand is finished, vehicle can be re-used for other passengers.

To build such a Cloud-Commuting system, we need to fundamentally understand urban trip demand patterns and evaluate the number of vehicles needed in the system, so as to serve all trip demands with good passenger quality-of-experience. In this paper, we try to make the first attempt to answer these critical questions by utilizing the real trajectory data from taxis in Shenzhen, China in 2014. Our main contributions are summarized as follows:

- Utilizing large-scale taxi trip data, we develop generative models to capture the arrival and service patterns of urban taxi trip demands.
- By modeling the Cloud-Commuting system as a queuing system, we propose an theoretical approach to estimate the total number of vehicles needed to serve all demands.
- Employing real-world trip demand data, we evaluate our proposed trip demand models and estimation approach, by comparing with empirical baseline models with promising results obtained.

The rest of the paper is organized as follows: Section II formally describes the motivation and defines the problem. Section III presents the framework of our model and provides detailed methodology. Section IV presents evaluation results over a large-scale taxi trajectory data. Related works are discussed in Section V, and the paper is concluded in Section VI.

II. MOTIVATION AND PROBLEM DEFINITION

In this section, we motivate our envisaged future Cloud-Commuting system, and define the problem of modeling the trip demand patterns and estimating the number of needed onservice vehicles.

A. Cloud-Commuting System

Existing urban transit systems suffer many well-known shortcomings. Public transits run along fixed routes with fixed time tables, and have limited coverage areas, while the private transits, such as taxis, are expensive and largely "on-demand", thus the supplies may not meet the demands in a timely manner. Taking taxis as an example, Fig 1 shows that on average more than 60% of taxis are idle over time.

To tackle these fundamental issues of the current urban transit systems, we envisage a forward-looking Cloud-Commuting based transit system for smart cities. In such system, every vehicle is controlled by centralized servers. When a trip demand is finished, vehicle can be re-used for other passengers. Employing giant pools of centralized taxis/shuttles, this

This work was supported in part by NSF CRII grant CNS-1657350, a research grant from Pitney Bowes Inc, NSF grants CNS-1411636, DTRA grant HDTRA1-14-1-0040, and DoD ARO MURI Award W911NF-12-1-0385. Guanxiong Liu and Menghai Pan contributed equally in this work.



Fig. 1: Idling taxis Fig. 2: Request pattern

transport system provides passengers with a fast, convenient, and low cost transport service. With the fast development of autonomous driving technology, it is promising to combine the self-driving vehicles with our envisaged Cloud-Commuting system.

However, to build a Cloud-Commuting system, we need to fundamentally understand urban trip demand patterns such as: (1) trip demand arrival pattern, (2) service pattern, (3) number of vehicles needed. In this paper, we try to make the first attempt to answer these critical questions by utilizing the real data from taxis in Shenzhen, China.

B. Problem Definition

Thanks to the fast development of location sensing technologies, the increasing prevalence of embedded sensors inside mobile devices, vehicles has led to an explosive increase of the scale of urban mobility datasets, including the trip demands data of passengers in urban areas.

Definition 1 (Trip demand). A trip demand of a passenger indicates the intent of a passenger to travel from a source location src to a destination location dst from a given starting time t_s with an expected trip duration Δt , which can be represented as a 4-tuple $\langle src, dst, t_s, \Delta t \rangle$.

Fig 2 shows the temporal distribution of urban taxi trip demands for each 10-minute time intervals in Shenzhen from 03/04/2014-03/06/2014, which exhibits a clear diurnal pattern. Such pattern is driven by the daily commuting needs between residential and working locations. Given such strong diurnal pattern, we divide each day into a few time intervals, and focus on the daily dynamics of trip demands over intervals.

Problem definition. Given a set of trip demands data, and T time intervals of a day, we aim to develop generative models to capture the patterns of urban trip demands, including (1) arrival rate distribution (2) service rate distribution, and introduce a model to estimate the total number of vehicles needed to serve all trip demands.

III. METHODOLOGY

In this section, we introduce our analytical framework of analyzing the urban trip demand pattern and estimating the bounding number of taxis to cover all trip demands.

A. Overview

Fig.3 illustrates our solution framework, that takes two sources of urban data as input and contains three key analytical stages: (1) trip demands extraction, (2) arrival and service



Fig. 3: Framework

pattern extraction, and (3) estimation of the bound of onservice taxis.

- Stage 1 (Trip demands extraction) Our taxi trajectory data are collected from each taxi. Each taxi trajectory consists of a sequence of time-stamped GPS points, where a GPS point is collected every 40 seconds on average. A GPS data point includes the time stamp, latitude, longitude, and binary indicator (indicating if a passenger is aboard). Moreover, the raw trajectory data are noisy in nature, with spatial errors from the actual locations, due to the accuracy limit of the GPS devices. Recall that each urban trip demand consists of four key elements: (1) starting location src, (2) ending location dst, (3) starting time t_s , (4) trip duration Δt . The first stage aims to clean and extract trip demands information from the raw trajectory data.
- Stage 2 (Arrival/Service pattern extraction) From the urban trip demands extracted from Stage 1, we can represent the urban trip arrival pattern as a sequence of trip starting time stamps, i.e., $\{t_{s1}, t_{s2}, ..., t_{sm}\}$, and find a generative model to explain the trip arrival pattern. Similarly, with all trip durations (as system service time), we will develop a model to characterize the service time pattern.
- Stage 3 (Bound of # on-service taxis) With generative models for arrival and service patterns of the urban trip demands, we can naturally view the taxi service system as a queuing system, with trip demands as the customers and taxis as the servers. In Stage 3, we employ queuing theory to develop an estimation method to quantify the bound of the number of taxis needed to serve all trip demands, which in turn explains how many taxis we can reduce in a Cloud-Commuting system, comparing to the current taxi service system.

B. Data Description

Our analytical framework takes two urban data sources as input, including (1) taxi trajectory data and (2) road map data. For consistency, both datasets are collected in Shenzhen, China in 2014. We introduce the details of these datasets below.



Fig. 4: Heat map of starting location

Fig. 5: Heat map of ending location

Taxi trajectory data are GPS records collected from taxis in Shenzhen, China during 2014. There were in total 17,877 taxis equipped with GPS sets, where each GPS set generates a GPS point every 40 seconds on average. Overall, a total of 51,485,760 GPS records are collected on each day, and each record contains five key data fields, including taxi ID, time stamp, passenger indicator, latitude and longitude. The passenger indicator field is a binary value, indicating if a passenger is aboard or not.

Туре	Counts	Туре	Counts
Motorway	563	Secondary	868
Trunk	258	Tertiary	1,393
Primary	745	Unclassified	16,829

TABLE I: Road Map Data in Shenzhen

Road map data. In our study, we use Google GeoCoding [1] to retrieve a bounding box of Shenzhen, which is defined between 22.45° to 22.70° in latitude and 113.75° to 114.30° in longitude. The covered area covers a total of $1,300km^2$. Within such a bounding region, we crawl road map data in Shenzhen from OpenStreetMap [2]. The road map data contain six levels of road segments, which are detailed in Table I and visualized in Fig.6.

C. Stage 1: Demands Extraction

In stage 1, we clean and extract the urban trip demands from the raw trajectory data.

Trajectory data cleaning. The trajectory data are noisy in nature. First of all, the GPS locations are with errors of around 15 meters. Secondly, there are GPS points outside the bounding box of Shenzhen. We conduct two steps to clean the noisy trajectory data, including map-matching and spatial filtering. *Map-matching* is a process that project the noisy GPS locations back to the road segments, which has been extensively studied in the literature We apply the map-matching technique [7] to our dataset. Secondly, we apply a simple *spatial filtering* step to remove GPS records that are outside the bounding region of Shenzhen.

Trip demand extraction. The passenger indicator field in the taxi trajectory data is the key enabler to extract the taxi trip demands. A taxi trip can be represented as a sequence of taxi GPS points with the passenger indicator as 1. The first and last GPS locations of the taxi trip capture the source/destination locations (*src*, *dst*) of a trip demand, and the corresponding time stamps characterize the trip starting/ending time t_s/t_e . The trip duration can be obtained as the elapsed time from t_s to t_d , i.e., $\Delta t = t_e - t_s$. Once we have all trip demand tuples

Fig. 6: Shenzhen road map

 $\langle src, dst, t_s, \Delta t \rangle$, we observe that there are a small number of trip demands with extremely short or long trip durations. From the size of the bounding region of Shenzhen and the road map, any trip could be done within 2 hours (including the rush hours with traffic congestion). Moreover, people would not take a taxi trip shorter than 2 minutes in general. Thus, we simply filter out those noisy taxi trips longer than 2 hours or shorter than 2 minutes, which may be due to the issues with hardware or data collection processes.

After the two steps, we obtain a total of 595,501 daily trip demands from our trajectory data. Fig.4 and Fig.5 show the geo-distributions of source and destination locations in Shenzhen during the morning rush hours 6–9AM on March 6th, 2014.

D. Stage 2: Arrival/Service pattern

The taxi service system in a city can be viewed as *a queuing system*. Each trip demand and the corresponding taxi trip represent a customer arrival event and a service event, respectively. Taxis are the servers in the system. To better understand the urban trip demands, now we characterize the arrival pattern and service pattern from the trip data, and model the maximum number of taxis needed in the queuing system. Arrival pattern is the distribution of the arrival events in a queuing system. In the taxi service system, each taxi demand is an arrival event. As defined below, the arrival rate captures the number of arrival events in a unit time.

Definition 2 (Arrival rate). *The arrival rate is the number of trip demands arriving the system within a unit time slot.*

We choose the time unit as one second, and count the number of arrived trip demands over each second in demand data we extracted from Stage 1. Fig.7 shows the distributions of the arrival rate in four different intervals of a day. The x-axis represent the arriving rates and the y-axis is the percentage of demands. The blue dots are obtained from original demands data, which nicely fit Poisson distributions. The green curves are the best fitting curves with Poisson distribution. The parameters λ 's of Poisson distributions are the mean arrival rates, which are listed in Table II for different time intervals in a day.

Service pattern. Our taxi trip data only capture the invehicle time of each trip. For simplicity, we consider the service time for a trip demand as the in-vehicle transit time ¹.

¹In fact, our results in this paper also work for a more comprehensive model with service time including passenger waiting time, in-vehicle transit time, and taxi return time. We omit this set of results due to the limited space



TABLE II: Parameters of arrival rate distributions

To extract the service time pattern from the demand data, we choose the unit time as minute. Fig.8 show the distributions of service time over four different time interval in a day. The x-axis represents the service time and the y-axis is the percentage of demands. The blue dots are from the raw demand data, which nicely follows Exponential distributions. The green curves are the best fitting curves of Exponential distributions. The green service rates which are the inverse of the mean service time, and the μ 's are listed in Table III for different time intervals in a day.

Time slot	12am-6am	6am-12pm	12pm-6pm	6pm-12am
μ	0.0607	0.0770	0.0736	0.0855

TABLE III: Parameters of service time distributions



E. Stage 3: Estimating the bound of number of on-service taxis

Now, we are in a position to introduce our queuing theory based approach to estimate the bound of number of on-service taxis in the taxi service system. We first introduce Theorem 1 below that explains the distribution of the number of on-service taxis in the system. The proof of Theorem 1 can be found in [3] on page 114.

Theorem 1. In an $M/G/\infty$ queuing system, let λ be the average arrival rate, and μ be the average service rate. then the number of customers (namely, on-service taxis) in system follows Poisson distribution with parameter ρ which is the utilization of the queuing system: $\rho = \frac{\lambda}{\mu}$.

To quantify the total number of needed taxis for serving all trip demands, we assume that there are a sufficient number of servers/taxis in the system, which means that the length of queue in our system is always empty. This way, each trip demand is guaranteed to be served by a taxi, when the demand arrives the system. The number of customers in system is same as the number of on-service servers (taxis). From Fig 7 and Fig 8, both of the distributions of arrival rates and service rates nicely follow Poisson distributions. Hence, we can denote our queuing model by $M/M/\infty$, which is a special case of $M/G/\infty$. According to Theorem 1, the Poisson distribution of on-service taxis can be obtained as $P(N = n) = \frac{\rho^n}{n!} e^{-\rho}$, with $\rho = \lambda/\mu$, and n as the number of on-service taxis.

Taking 12pm-6pm slot as an example, the average arrival rate is $\lambda = 8.4415$ and the average service rate is $\mu = 2.2093$, then the utility $\rho = 6877.55061$, which captures the mean of the number of on-service taxis in the system. (Note that the units of λ and μ are consistent, when calculating ρ). Then the probability distribution and cumulative distribution of the number of on-service taxis is showed in Fig.9. We choose a probability threshold of 99.99% to obtain the theoretical bound of the maximum number of on-service taxis. From Fig.9b, we observe that a maximum of 7,202 taxis are on service for 99.99% of time units.

Time slot	12am-6am	6am-12pm	12pm-6pm	6pm-12am
mean	4089	5934	6878	6319
maximum	4325	6116	7202	6612

TABLE IV: Theoretical mean and max of 4 time slots

IV. EVALUATION

In this section, we use real taxi trip data to conduct experiments to evaluate how well our proposed models of arrival and service distributions characterize the trip demand patterns from real data, and how accurate our bound estimation approach can capture the ground-truth bound of the on-service taxis.

A. Evaluation settings

Time intervals in a day. We observe that the trip demand arrival and service rates change dramatically over time intervals in a day. In our evaluations, we divide a day into different numbers of time intervals, and evaluate how the granularities affect the performances of our proposes models. We choose different numbers of equal-length intervals in a day, ranging within [1,4,8]. For example, when dividing a day into 4 intervals, we have the cutting-off times as [12am, 6am, 12pm, 6pm]. Similarly, we have cuttingoff times as [12am, 3am, 6am, 9am, 12pm, 3pm, 6pm, 9pm]when 8 time intervals are divided. We will show shortly that within each time interval, the arrival and service distributions stay relatively stable.

Baseline. We compare the bound of number on-service taxis obtained by our approach with an empirical bound, obtained from a simulation process as follows: For every second, we count the number of on-service taxis from the data. An empirical bound on the on-service taxis of a time interval, e.g., 12pm-6am, can be obtained as the maximum number of on-service taxis per second.

Metrics. From our approach, we can obtain the theoretical maximum and mean number of on-service taxis, we compare these values with those obtained from the empirical approach. We use the maximum gap between the theoretical and empirical maximums and the mean gap between the theoretical and empirical means to evaluate how accurate our proposed models capture the trip demand patterns.

B. Evaluation Results

1) 4 time intervals Results: Fig.10(a) shows the theoretical and the empirical maximums during the four time intervals. In slots 12pm-6pm and 6pm-12am, the theoretical and empirical maximums match each other, where within 12am-6am and 6am-12pm, the gaps between the two maximums are relatively large. The reason is that numbers of on-service taxis during 12am-6am and 6am-12pm fluctuate more, say, with higher variance.

Fig.10(b) shows the theoretical and the empirical mean of 4 time intervals. We can see that the two means match each other very well.





Fig. 10: Comparison with 4 Time Slots



Fig. 12: Comparison with 8 Time Slots

2) 8 time intervals Results: When we divide a day into 8 time intervals, the arrival and service patterns of 8 time slots still follow Poisson distribution and Exponential distribution as shown in Fig.11, which shows the arrival rate distribution and the service time distribution within the time interval of 12pm-3pm. Fig.12(a) shows the theoretical and the empirical maximum of 8 time intervals. Overall, the theoretical and empirical maximums match each other better than those with 4 time intervals. The reason is that by dividing a day into 8 intervals, the variance within each interval decreases. Fig.12(b) shows the theoretical and the empirical means of 8 time slots. We can see that the two means match each other well.

3) Estimation Gap over Number of Intervals: Fig.13 shows relations between max/mean gaps and number of time slots. One interesting phenomenon is that with increasing number of time intervals, the max gap becomes smaller, while the mean gap becomes larger. This is because means are calculated among all data in each time interval, while max is always dominated by the extreme data point with the highest onservice taxi number. When we increase the number of time slots, the amount of data in each time interval decreases, thus



the variance of the estimated mean increases. On the other hand, with a smaller time interval, the variance of estimated max decreases.

V. RLATED WORK

To the best of our knowledge, we are the first to build a dynamic model in helping people understand urban transit needs by capturing several key features from real world trip demands. In this section, we will talk two topics that are closely related to our work which include (1) mobility-on-demand system (2) urban computing.

Queuing Theory analysis on mobility-on-demand system (MoD). MoD is an emerging concept in solving urban transportation problems like unbalanced supply-demand rates and traffic congestion. As an urban transportation system, MoD aims at providing transit supplies, such as shuttle/taxi services according to dynamic urban trip demands. In [11], authors design a simulation platform, which could be used to explore the performance of autonomous vehicle based MoD system, under various operation models. In another work [4], a general mathematical model is proposed, which could make real-time assignment decision in high-capacity ride-sharing system. This model is designed to handle a large number of passenger demands and dynamically generate optimal assignment solution to urban trip demands. In [15] and [13], authors propose two spatial queueing-theoretical models, that capture salient dynamic and stochastic features of customer demand, for Autonomous mobility-on-demand system which has autonomous vehicles in it. However, given the promising vision of MoD, there is a fundamental research gap before developing system implementation and design plans, which is to understand the underlying patterns of urban trip demands and estimating the capacity needs of an MoD system. In this paper, we bridge this gap by developing generative models to capture and explain urban trip arrival and service patterns, and design a estimation model to quantify the system capacity needs of MoD-like system.

Urban Computing is a relatively new and thriving research area which integrates urban sensing, data management and data analytic together as a unified process to explore, analyze and solve existing critical problems related to people's every-day life [8], [5], [10], [14], [6], [9]. The first kind of topics in urban computing we want to talk about is related to electricity vehicle. As we can see from papers [8], the trajectory data from electrical taxis is collected and a data driven solution

is proposed. This solution reorganize assignment plan of charging station and number of facilities based on historical trajectory of electrical vehicles to optimize the average driving distance in finding a charging station. [14] develops novel models to predict future crowd flow traffic in subway stations. Different from these works, in this paper, we develop quantitative models that characterize the urban trip demand patterns, which shed lights on how to design smarter future urban transportation system, with better passenger quality-ofexperience (QoE).

VI. CONCLUSION

In this paper, we envisage a Cloud-Commuting transportation system, that employ a giant pool of centralized taxis/shuttles to better cope with the dynamic urban trip demands. To better understand the feasibility of such a system, we develop generative models to capture fundamental demand arrival and service patterns, and introduce a novel model to estimate the total number of vehicles needed to serve all urban demands. Experiments conducted using real taxi data demonstrated that our proposed models can precisely capture the underlying urban trip demand pattern, and provide an accurate estimate of the system capacity for the envisaged Cloud-Commuting system.

REFERENCES

- Google GeoCoding. https://developers.google.com/maps/documentation/ geocoding/.
- [2] OpenStreetMap. http://www.openstreetmap.org/.
- [3] I. Adan and J. Resing. Queueing theory, Eindhoven University of Technology, 2010.
- [4] J. Alonso-Mora, S. Samaranayake, A. Wallar, E. Frazzoli, and D. Rus. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, 2017.
- [5] Y. Ding, Y. Li, K. Deng, H. Tan, M. Yuan, and L. M. Ni. Dissecting regional weather-traffic sensitivity throughout a city. In *ICDM*, 2015.
- [6] Y. Ding, Y. Li, K. Deng, H. Tan, M. Yuan, and L. M. Ni. Detecting and analyzing urban regions with high impact of weather change on transport. *IEEE Transactions on Big Data*, 2016.
- [7] H.-P. Hsieh, S.-D. Lin, and Y. Zheng. Inferring air quality for station location recommendation based on urban big data. In SIGKDD, 2015.
- [8] Y. Li, J. Luo, C.-Y. Chow, K.-L. Chan, Y. Ding, and F. Zhang. Growing the charging station network for electric vehicles with trajectory data analytics. In *ICDE*, 2015.
- [9] C. Liu, K. Deng, C. Li, J. Li, Y. Li, and J. Luo. The optimal distribution of electric-vehicle chargers across a city. In *ICDM*. IEEE, 2016.
- [10] B. Lyu, S. Li, Y. Li, J. Fu, A. C. Trapp, H. Xie, and Y. Liao. Scalable user assignment in power grids: a data driven approach. In *SIGSPATIAL GIS*. ACM, 2016.
- [11] K. A. Marczuk, H. S. S. Hong, C. M. L. Azevedo, M. Adnan, S. D. Pendleton, E. Frazzoli, et al. Autonomous mobility on demand in simmobility: Case study of the central business district in singapore. In CIS-RAM 2015.
- [12] D. of Economic and S. Affairs. World urbanization prospects 2014. https://esa.un.org/unpd/wup/Publications/Files/WUP2014-Highlights. pdf.
- [13] M. Pavone. Autonomous mobility-on-demand systems for future urban mobility. In Autonomous Driving, pages 387–404. Springer, 2016.
- [14] E. Toto, E. A. Rundensteiner, Y. Li, R. Jordan, M. Ishutkina, K. Claypool, J. Luo, and F. Zhang. Pulse: A real time system for crowd flow prediction at metropolitan subway stations. In *ECML-PKDD*. Springer, 2016.
- [15] R. Zhang, K. Spieser, E. Frazzoli, and M. Pavone. Models, algorithms, and evaluation for autonomous mobility-on-demand systems. In *American Control Conference (ACC)*, 2015, pages 2573–2587. IEEE, 2015.