Data-Driven Digital Advertising with Uncertain Demand Model in Metro Networks

Ruobing Jiang[†], Zhenni Feng[†], Desheng Zhang [♯*], Shuai Wang[§], Yanmin Zhu[†], Fan Zhang[‡], Tian He[†§]
[†]Shanghai Jiao Tong University, China, [♯]Rutgers University, USA,
[§]University of Minnesota, USA, [‡]SIAT, China

Abstract—Nowadays most metro advertising systems schedule advertising slots on digital advertising screens to achieve the maximum exposure to passengers by exploring passenger demand models. However, our empirical results show that these passenger demand models experience uncertainty at fine temporal granularity (e.g., per min). As a result, for fine-grained advertisements (shorter than one minute), a scheduling based on these demand models cannot achieve the maximum advertisement exposure. To address this issue, we propose an online advertising approach, called FineUDM, based on the uncertain passenger demand modeling for both entering passengers and exiting passengers. FineUDM combines coarse-grained statistical demand modeling and fine-grained real-time demand modeling by leveraging historical passenger demands, real-time card-swiping records, and passenger mobility patterns. Based on this uncertain demand model, it schedules advertising time online based on robust receding horizon control to maximize the advertisement exposure. We evaluate the proposed approach based on an one-month sample from our 530 GB real-world metro fare dataset with 16 million cards. The results show that our approach provides a 61.5% lower traffic prediction error and 20% improvement on advertising efficiency on average.

Index Terms—Digital Advertising, Uncertain Passenger Demand, Data-driven Traffic Volume Prediction.

1 Introduction

Digital advertising systems [1] in metro networks obtain increasing preference by advertisers because of the extensive exposure to a large number of passengers, e.g., in New York City, the annual ridership is nearly 1.8 billion [2] according to the Metropolitan Transportation Authority. In such a metro advertising system, the key design issue is how to schedule given advertising time (rented by advertisers) on a digital screen to maximize passing audience for the maximum advertisement exposure during all the advertising time, which is a key indicator of the potential advertising revenues [3].

Historically, an intuitive approach for this issue is to preferentially display advertisements in those time periods with more historical passengers, i.e., a passenger demand model based on historical passenger data collected in the automatic fare collection (AFC) system [4], [5]. The assumption behind such an approach with a historical-data-based demand model is that the daily distribution of passenger traffic volumes, i.e., passenger demand, in the same metro station is fairly predictable with historical distributions. Thus, higher historical passenger demand in a past time period indicates higher future passenger demand for the same time period in the future day.

However, our empirical study in the metro network of Shenzhen (the twin city of Hong Kong and has more than 11 million residents) clearly shows that only coarse-grained (e.g., one hour) passenger demand is predictable based on historical data, but fine-grained (e.g., one minute) passenger demand is with uncertainty and thus unpredictable based on historical data. Further, we argue that the fine-grained (instead of coarse-grained) passenger demand model is essential to advertising time scheduling, because an advertisement is normally in a fine-grained length (e.g., less than one minute). As a result, we face a key challenge to predict fine-grained passenger demand to effectively schedule advertising time.

To address this challenge, we propose an online advertisement scheduling approach, named *FineUDM*, based on robust receding horizon control with an uncertain demand model. Specifically, the uncertain demand model in a given metro station during a given time period contains two independent parts of passenger demand, namely, entering passenger demand and exiting passenger demand, which has been studied in our previous work [6].

1) Entering passenger demand modeling: We introduce RealPoisson, a non-homogeneous Poisson model with a real-time tuner to predict the fine-grained real-time entering passenger demand. RealPoisson, instead of using a classic homogeneous Poisson model, adopts time-varying rate parameter for entering passenger demand modeling because the entering demand in a station varies from hour to hour, e.g., the entering demand during peak hours will be much higher than that during off-peak hours.

2) Exiting passenger demand modeling: We combines coarse-grained statistical demand modeling and fine-grained Bayesian demand modeling by inferring individual arrivals using both real-time and historical, *instead of only historical*, data in the AFC system. This is because given real-time entering stations and time, we can accurately infer destinations and arriving time for passengers based on the low conditional entropy of passenger destinations and

* *Desheng Zhang is the corresponding author.*

predictable travel durations in the metro network, learned from our extensive empirical study.

In addition, we design an advertising scheduling algorithm based on receding horizon control to maximize advertisement exposure under our uncertain demand model. Specifically, our main contributions are as follows:

- To our knowledge, we perform the first work, named FineUDM, to optimize advertisements exposure with large-scale data from the metro automatic fare collection system. We already released the sample data for the benefit of research community.
- We design a two-level uncertain demand model based on our extensive empirical study, which reveals that the coarse-grained passenger demand is predictable with the historical data while the fine-grained passenger demand is with uncertainty and thus unpredictable based on history. Thus, we combine coarse-grained statistical demand modeling and fine-grained real-time demand modeling to improve the demand prediction performance.
- We propose RealPoisson, a non-homogeneous Poisson model with a real-time tuner to model uncertain fine-grained entering demand. The two challenging designs of RealPoisson include the time-varying rate parameter based on historical data and the demand trend tuner based on real-time traffic. The former removes the fix-rate limitation of a classic homogeneous Poisson model and exploits the inherent difference among hourly entering traffic for each metro station according to history. The latter restores the averaged randomness for each fine-grained entering demand based on near-historical demands.
- We propose EveryoneCounts, a bayesian model to predict uncertain fine-grained exiting demand by leveraging (i) the real-time card swiping records collected by the AFC system, (ii) the low conditional entropy of passenger destinations, and (iii) the predictable travel time between different stations.
- We propose an online digital advertising approach based on robust receding horizon control, which exploits our passenger demand model to allocate coarse-grained advertising time offline but adjust fine-grained advertising time online to maximize advertisement exposure.
- We evaluate the performance of our approach through extensive data-driven evaluation based on one-month sample from our 530 GB real-world metro fare dataset with 16 million cards in Shenzhen, China. Compared to statistical approaches, the proposed approach has a 61.5% lower prediction error of fine-grained traffic volumes, leading to a 20% improvement in advertising efficiency.

The rest of the paper is organized as follows. Section 2 defines the advertising optimization. Novel empirical results are presented in Section 3. Sections 4-9 show the detailed design and its performance. Section 10 discusses the real-world issues. We introduce the related work in Section 11 and conclude the paper in Section 12.

2 Models and Problem Definition

In this section, we first present the scenarios and the advertising models. Then, we provide the formal definition of the advertising efficiency optimization problem. The mainly used notations are summarized in Table I.

2.1 Scenarios and Models

We focus on the metro stations where passengers need to swipe their metro cards at the entrance/exit to enter/exit a station. We model the metro network as a set of connected metro stations, denoted by $\Psi = \{\psi_1, \cdots, \psi_d, \cdots, \psi_m\}$. As shown in Fig. 1, we show the metro network of Shenzhen and a scenario where digital screens are installed and exposed to passengers between AFC machines and platforms of a station. The lighter the icon, the higher the average passenger demand.

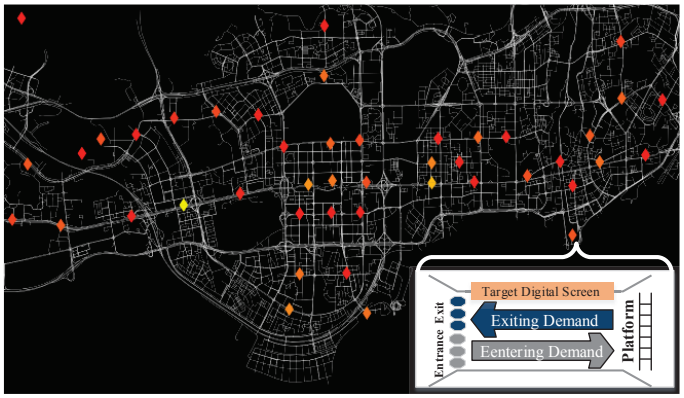


Figure 1: An illustration of the digital screen advertising in Shenzhen metro network: both the entering passengers (from outside of the metro network) and exiting passengers (from other stations within the metro network) will pass by the target screen when they enter or exit the station.

The metro network sells advertising time of each digital screen in each station to advertisers and charges them based on the length of advertising time. The maximum length of daily advertising time for each digital screen at all stations, denoted by T (e.g., 24 hours), is divided into small time slots with equal length τ (e.g., one minute). Suppose γ advertisers buy advertising time on digital screens and the lengths of their advertising time are α_d^i, where $1 \leq i \leq \gamma$ and $1 \leq d \leq m$. We consider all advertisers with same priority, although our method can also be used for multiple priorities. For a given slot t_j and a station ψ_d, the passenger demand $\beta_{j \cdot d}$ is given by the *slot traffic volume*, i.e., the number of passengers passing by the digital screen during t_j in station ψ_d. Note $\beta_{j \cdot d} = \beta_{j \cdot d}^1 + \beta_{j \cdot d}^0$ where $\beta_{j \cdot d}^1$ is the slot traffic volume of passengers entering station ψ_d and $\beta_{j \cdot d}^0$ is the exiting slot traffic volume.

2.2 Problem Definition

Given the advertisement utility (i.e., T) of metro stations, the objective of the metro advertising system is to provide the best advertising service to the advertisers in terms of optimized advertising efficiency. With the above models and

TABLE 1: Main Notations Used in This Paper

Notation	Description
T	The length of a time unit, e.g., 24 h, at a station ψ_d
τ	The length of time slots, e.g., 3 min
α_d^i	The length of advertising time for a station ψ_d and advertiser i
$\beta_{j\cdot d}$	The traffic volume of jth slot at a station ψ_d
$\rho_{j\cdot d}^i$	The AD schedule for jth slot at a station ψ_d for advertiser i

notations, we now formally define a key term – advertising efficiency and its corresponding optimization problem.

Definition 1 (**Advertising Efficiency**). *The traffic volume over the total advertising time on the screens at all stations, i.e.,*

$$\frac{\sum_{i=1}^{\gamma}\sum_{d=1}^{m}\sum_{j=1}^{\frac{T}{\tau}}\beta_{j\cdot d}\times\rho_{j\cdot d}^i}{\sum_{i=1}^{\gamma}\sum_{d=1}^{m}\alpha_d^i}, \tag{1}$$

where the schedule of the jth slot at the station ψ_d for the advertiser i, $\rho_{j\cdot d}^i$, is given by,

$$\rho_{j\cdot d}^i=\begin{cases}1, \text{ if } t_j \text{ is an AD slot for advertiser } i \text{ at station } \psi_d\\ 0, \text{ otherwise}\end{cases}. \tag{2}$$

For the numerator, the innermost summation is for all AD slots in a particular station ψ_d for a particular advertiser; the middle summation is for all AD slots in all stations for a particular advertiser; the outmost summation is for all AD slots in all stations for all advertisers. For the denominator, the summation is the total advertising time.

Definition 2 (**Advertising Efficiency Optimization**). *Given the total advertising time and the slot length, a set of schedules $\rho_{j\cdot d}^i$ needs to be made, so that the advertising efficiency is maximized, i.e.,*

$$\max_{\rho_{j\cdot d}^i}\frac{\sum_{i=1}^{\gamma}\sum_{d=1}^{m}\sum_{j=1}^{\frac{T}{\tau}}\beta_{j\cdot d}\times\rho_{j\cdot d}^i}{\sum_{i=1}^{\gamma}\sum_{d=1}^{m}\alpha_d^i},$$
$$s.t.\ \tau\times\sum_{j=1}^{\frac{T}{\tau}}\rho_{j\cdot d}^i=\alpha_d^i,\forall i\in[1,\gamma],\forall d\in[1,m]. \tag{3}$$

To solve this optimization problem, we need to find a schedule $\rho_{j\cdot d}^i$ for different AD slots in different stations for different advertisers. In our setting, γ, m, T, τ, and α_d^i where $\forall i\in[1,\gamma]$ are given in advance. But the passenger demand $\beta_{j\cdot d}$ for a particular slot in a particular station has to be obtained by a demand model. Such passenger demand includes passengers who enter or exit the target station. Generally speaking, passengers that enter and leave through the station are independent, we model them with two separate fine-grained models and add them together as the entire uncertain demand model in the paper.

In this work, as shown in Section 3, we found that fine-grained passenger demand models experience uncertainty. Thus, we formulate the passenger demand $\beta_{j\cdot d}$ in our optimization as a variable, instead of a fixed value. In this work, we assume $\beta_{j\cdot d}$ belongs to some uncertainty set where $\underline{\beta}_{j\cdot d}\le\beta_{j\cdot d}\le\overline{\beta}_{j\cdot d}$. As a result, we formulate a robust optimization problem as follows.

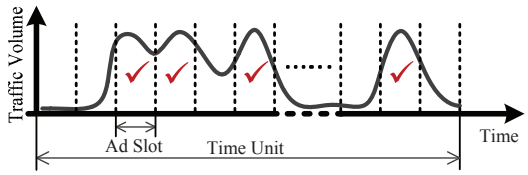


Figure 2: The illustration to explain the definition of the advertising efficiency optimization problem.

Definition 3 (**Robust Advertising Efficiency Optimization**).

$$\max_{\rho_{j\cdot d}^i}\min_{\beta_{j\cdot d}}\frac{\sum_{i=1}^{\gamma}\sum_{d=1}^{m}\sum_{j=1}^{\frac{T}{\tau}}\beta_{j\cdot d}\times\rho_{j\cdot d}^i}{\sum_{d=1}^{m}\sum_{i=1}^{\gamma}\alpha_d^i}$$
$$s.t.\ \tau\times\sum_{j=1}^{\frac{T}{\tau}}\rho_{j\cdot d}^i=\alpha_d^i,\forall i\in[1,\gamma],\forall d\in[1,m]$$
$$\beta_{j\cdot d}\in[\underline{\beta}_{j\cdot d},\overline{\beta}_{j\cdot d}],\forall j\in[1,\frac{T}{\tau}],\forall d\in[1,m] \tag{4}$$

In this problem, the key challenge is how to obtain an uncertainty set of passenger demand, i.e., $\underline{\beta}_{j\cdot d}$ and $\overline{\beta}_{j\cdot d}$ and then to determine the schedule $\{\rho_j\}_{j=1}^{\frac{T}{\tau}}$ with an online fashion, so that the best $n=\frac{\sum_{i=1}^{\gamma}\alpha_d^i}{\tau}$ slots with highest passenger demand can be selected as the advertising slots for all the advertisers in the given station d. As illustrated by Fig. 2, if the uncertain passenger demand for each time slot can be predicted, the best n slots with highest passenger demand will be selected as AD slots. Then, all the selected AD slots can be allocated to all the advertisers (assumed to have same priority in this work) according to their priorities. Note that the allocation strategy is not the focus of this work.

As follows, we conduct an empirical study in Section 3 before we present our demand model and scheduling.

3 Empirical Study

To predict future passenger demand, we conduct an extensive empirical study on the real-world AFC records collected in the metro system in Shenzhen, China.

3.1 Dataset

In this paper, we utilize two sample datasets from a streaming dataset of smartcard transactions in the metro AFC system of Shenzhen, which is the twin city of Hong Kong and has more than 11 million residents. Each card swiping record includes card ID, device ID, swiping in or out, date, time as well as metro station ID among 118 metro stations. In this paper, two sample datasets, named as Dataset A and Dataset B, are used. The summary of Dataset A and Dataset B as well as the format of each data record are illustrated in Fig. 3. As shown, Dataset A contains 2,807,973 smart cards, and the records range from Dec 12, 2013 to Dec 25, 2013. The average daily number of card swipings in the metro system is more than 3 million. The average daily traffic per station is about 25,000. Dataset B contains the streaming card swiping data of 16,000,000 passengers from July 1, 2011, including smartcards used in both metro systems and bus systems in Shenzhen.

	Dataset A	Dataset B
Collection Period	13/12/12- 13/12/25	11/07/01- Now
Number of Cards	2,807,973	16,000,000
Number of Records	41,270,178	6,212,660,742
Format: Card ID, Station ID, Device ID, Date &Time, Swipe in/out		

Figure 3: The dataset summary and the record format.

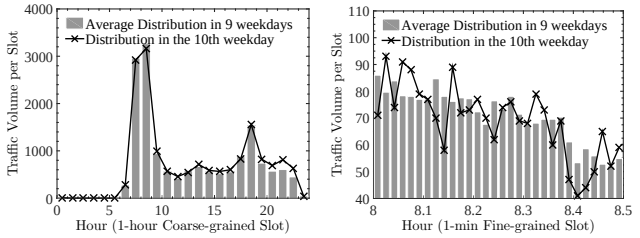
We utilize Dataset A as a sample to study all metro stations during a two week period. Dataset B is used for the large-scale evaluation. In the rest of this section, although similar observations are found throughout all the 118 stations, we focus on our empirical study results over the Xixiang station, which is a representative commute station.

In the following, we offer several observations on both entering demand and exiting demand at different levels of temporal granularities. In Subsection 3.2, we first investigate traffic volumes of entering demand and offer a few interesting observations. We further explore exiting demand In Subsection 3.3 and get similar observations with entering demands.

3.2 Observations on Entering Demand

3.2.1 Certainty of Coarse-grained Entering Demand

We first study the trend of coarse-grained entering demand of passengers in terms of traffic volumes as shown in Fig. 4(a), which plots the trends of a weekday (polyline) and the nearby historical nine weekdays (bars). The temporal granularity of the traffic volume is one hour, which is coarse-grained compared to the typical duration of an advertisement, e.g., one minute.



(a) The distribution of the coarse-grained entering demand in a weekday and the average distribution of its near historical weekdays.

(b) The distribution of the fine-grained entering demand in the rush hour (08:00 am to 08:30 am) in a weekday and its near historical weekdays.

Figure 4: Comparison of coarse-grained and fine-grained entering demand.

Observation 1.1. *Predictable* and *Smooth* Coarse-grained Entering Demand: The trend of the single weekday highly matches the trend of its near historical weekdays. The curve is smooth in general, except for the rush hour. In other words, the trend of the coarse-grained entering passenger demand in a station is predictable based on the nearby history records.

The main reason for the coarse-grained predictability and smoothness is that the entering passenger demand is relatively stable during near days, in a specified area around a station. Based on Observation 1.1, we conclude that the distribution of coarse-grained entering passenger in a station can be accurately predicted using the average history distribution.

3.2.2 Uncertainty of Fine-grained Entering Demand

We now introduce the trend of the fine-grained entering passenger demand. Fig. 4(b) plots the distribution of fine-grained entering demand during the rush hour (from 08:00 am to 08:30 am) of a weekday and its nearby historical nine weekdays.

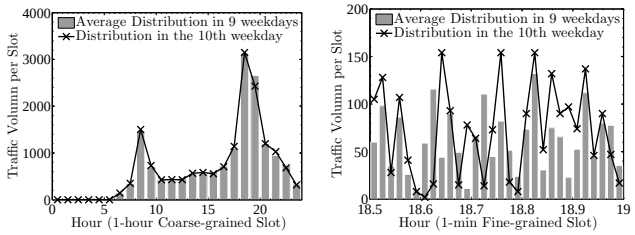
Observation 1.2. *Unpredictable* and *Fluctuated* Fine-grained Entering Demand: On the one hand, the trend of the fine-grained passenger volume entering the station in a weekday does not follow the average over nearby historical weekdays. That means the traffic volume in the same fine-grained slot largely varies in different days and thus unpredictable according to purely historical records. On the other hand, the trend of the fine-grained passenger volume, even during rush hours with crowded passengers, is fluctuated with continuous peaks and valleys. For example, some valleys next to a peak even have a traffic volume as low as the average volume in slack hours.

The unpredictability and fluctuation of the fine-grained entering volume is resulted by the randomness of human mobility. It is common for a commuter to enter the same station at different minute in the morning for different weekdays even when the commuter needs to arrive at the workplace at the same minute everyday.

3.3 Observations on Exiting Demand

3.3.1 Certainty of Coarse-grained Exiting Demand

We then study the trend of coarse-grained exiting demand of passengers in terms of traffic volumes as shown in Fig. 5(a), which plots the trends of a weekday (polyline) and the average over nearby historical nine weekdays (bars). The temporal granularity of the traffic volume is also one hour.



(a) The distribution of the coarse-grained exiting demand in a weekday and the average distribution of its near historical weekdays.

(b) The distribution of the fine-grained exiting demand in the rush hour (06:30 pm to 07:00 pm) in a weekday and its near historical weekdays.

Figure 5: Comparison of coarse-grained and fine-grained exiting demand.

Observation 2.1. *Predictable* and *Smooth* Coarse-grained Exiting Demand: Similar with the coarse-grained entering demand, the trend of the coarse-grained exiting demand in a single weekday highly matches the average trend of the near historical weekdays. We can also make the conclusion that the trend of the coarse-grained exiting demand in a given station is predictable based on the nearby history.

It is worth noting that the rush hours for the entering demand and the exiting demand in this given station occur at different time of a day. We can clearly find in Fig. 5(a) the entering rush hours occur at 08:00am and in Fig. 4(a) the exiting rush hours occur at 07:00pm. This is because Xixiang is a station located in a residential area and commuters usually go to work from the station (i.e., entering demand) in the morning and return home (i.e., exiting demand) in the afternoon.

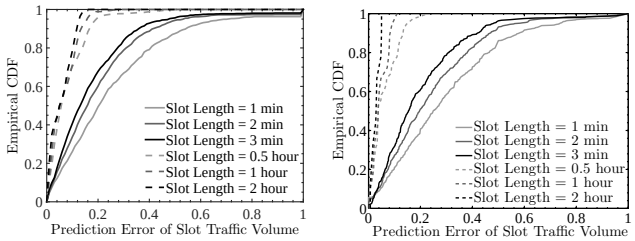
3.3.2 Uncertainty of Fine-grained Exiting Demand

For the fine-grained exiting demand, we plot in Fig. 5(b) the distribution of the demand during the rush hour (from 06:30 pm to 07:00 pm) of a weekday and the average over nearby historical nine weekdays.

Observation 2.2. *Unpredictable* and *Fluctuated* Fine-grained Exiting Demand: Compared with the entering demand, the trend of the fine-grained exiting demand is even more random and more fluctuated.

The fine-grained exiting demand is unpredictable due to both the randomness of human mobility and the deviation of the arrival time for the same train in different days. Moreover, different from entering demand where the commuters need to arrive at the workplace on time everyday, the exiting demand is more flexible since the returning commuters are not required to arrive at home at the same time everyday. For the greater fluctuation of fine-grained exiting demand, the reason is that the exiting passengers are highly limited by the schedule of metro trains. Between the arrivals of two successive trains in a metro station, few passengers exit the station.

Inspired by Observation 1.2 and Observation 2.2, we aim to propose fine-grained advertising instead of coarse-grained advertising because the existence of valley slots during a coarse-grained period (e.g., rush hours) pulls down the advertising efficiency of that period. Only the fine-grained selection of all the high-volume slots leads to better advertising efficiency.



(a) The CDFs of the prediction error of ENTERING volumes.

(b) The CDFs of the prediction error of EXITING volumes.

Figure 6: The CDFs of the prediction error of both coarse-grained and fine-grained volumes in a weekday based on the near history.

3.4 Historical Prediction Performance

We finally reveal the different performance of the historical prediction of fine-grained and coarse-grained passenger demand. The historical prediction error of both entering and exiting passenger demands with varying lengths of time slots (i.e., 1 min, 2 min, 3 min, 0.5 h, 1 h, and 2 h) are shown in Fig. 6(a) and Fig. 6(b), respectively. The demand prediction error for time slot t_j, denoted by δ_j, is computed as $\delta_j = \frac{|\beta_j - \hat{\beta}_j|}{\max\{\beta_j, \hat{\beta}_j\}}$, where β_j is the true passenger volume of time slot t_j and $\hat{\beta}_j$ is the average historical traffic volume. Specifically in this study, β_j is the slot volume in one weekday and $\hat{\beta}_j$ is the average slot volume in near historical nine weekdays.

From the figures, we have the following observations on the prediction performance of pure historical data. On the one hand, the prediction error of coarse-grained traffic volume for both entering passengers and exiting passengers is much lower than that of fine-grained traffic volume. For example in Fig. 6(b), when the slot length is 1 h, the exiting traffic prediction errors of around 99% slots are lower than 0.1. However, more than 40% slots have a prediction error larger than 0.3 when the slot length is 1 min. On the other hand, we can find that the fine-grained prediction error of exiting volumes is a little higher than that of entering volumes. For example in Fig. 6(a)(entering), when the slot length is also 1 min, there are only about 30% slots (more than 40% for exiting volume) have a prediction error larger than 0.3. This observation confirms with the difference between Fig. 4(b) and Fig. 5(b), i.e., the fine-grained exiting demand is even more random and fluctuated than the fine-grained entering demand (Observation 2.2).

3.5 Summary

Based on the empirical study, we have following important conclusions: (i) According to Observation 1.1 and 2.1, the coarse-grained passenger demand is predictable based on historical distributions, which provides us a global view on the coarse-grained passenger demand in a future day. (ii) According to Observation 1.2 and 2.2, the fine-grained passenger demand is unpredictable based on historical distributions. As a result, we propose an uncertain passenger demand model to predict fine-grained entering demand and exiting demand based on a special kind of Poisson process and prediction of individual arrivals, respectively. (iii) According to Observation 1.2 and 2.2, the fine-grained, instead of coarse-grained, advertising scheduling should be applied due to fluctuated fine-grained passenger demands. Thus, the selection of peak fine-grained slots among all the fluctuated slots is enabled to achieve higher advertising efficiency. So we design an online advertising scheduling based on robust receding horizon control. As follows, we first give an overview of our approach in Section 4, and then introduce our demand model and AD scheduling at Section 6, 7, and 8, respectively.

4 Methodology Overview

Inspired by the empirical study, we propose the "Everyone Counts" design to improve the advertising efficiency by robust receding horizon control based on a two-level passenger demand model. As follows, we introduce the overview of the proposed approach. In the rest of the paper, we use **frame**, denoted by $\{f_l\}_{l=1}^{\frac{T}{h}}$, to represent the coarse-grained time slot whose length, denoted by h, is in hour level. We use **slot**, with minute-level length τ, denoted by

$\{t_j^l\}_{j=1}^{\frac{h}{\tau}}$, to represent the jth fine-grained slot in frame f_l. Our demand modeling and AD scheduling are based on these two temporal units.

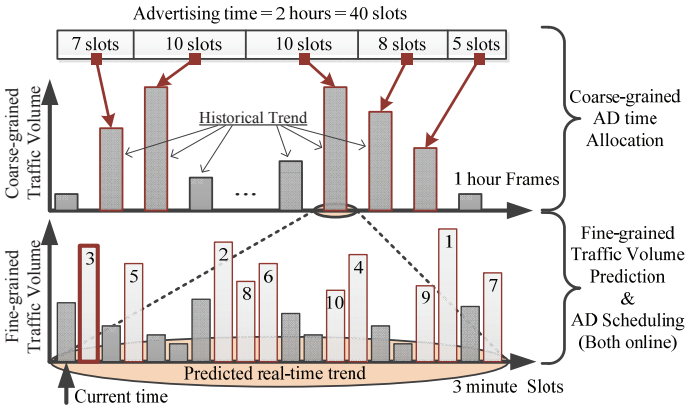


Figure 7: An example illustrating the overview of the proposed approach.

Given the fixed advertising time α_d^i and slot length τ, the proposed approach (i) allocates the total $n = \alpha_d^i/\tau$ AD slots to coarse-grained frames based on our *coarse-grained demand model* obtained by historical AFC data, and (ii) allocates all AD slots in a particular frame based on our *fine-grained demand model* obtained by real-time AFC data plus individual mobility patterns.

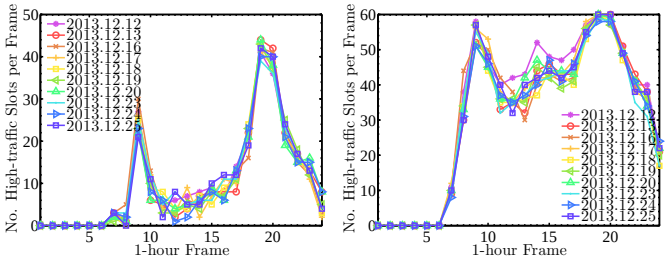
We use an example regarding to a specific metro station ψ_d and a specific advertiser i to illustrate the main process of the proposed approach, since scheduling at different stations is independent among each other. As shown in Fig. 7, the length of the advertising time α_d^i is two hours, and the length of slots τ is 3 minutes. Then the total number of advertising slots (AD slots) n is 40. The length of frames h is 1 h. Thus, there are 24 frames in the day and 20 slots in each frame. Our scheduling has two key steps. (i) According to the historical distribution of coarse-grained passenger demand in frames, n AD slots are allocated to $\{f_l\}_{l=1}^{24}$. Note the allocation based on the global view of high-traffic slots distribution in frames enables frame-scale (instead of day-scale) passenger demand prediction, which leads to higher demand prediction accuracy. (ii) Taking one of the frames with 10 AD slots for example, fine-grained passenger demand (for both entering passengers and exiting passengers) in this frame is predicted online and updated in each time slot with a non-homogeneous Poisson model and a receding horizon of newly occurred AFC records. Suppose the best 10 slots with high traffic volumes are those illustrated in the figure, according to the prediction at the current time (i.e., the first slot in the frame). These best 10 slots may be replaced by other slots in further predictions with a receding horizon. Based on this slot-level demand model, an online AD schedule is made for all the slots in the current frame, and the schedule is implemented only for the nearest future slot, i.e., the next slot of the current one. In this example, the next slot ranks 3rd among all the slots in terms of traffic volume, which will be selected as an AD slot. The online prediction and scheduling continue until all the slots in the frame are scheduled.

We next present the detailed design of our two-level demand model and AD scheduling in the following four sections.

5 Coarse-grained Demand Model

Given historical AFC data, our coarse-grained passenger demand model is a statistical model, i.e., at frame-level using the historical average passenger demand as the future passenger demand. Based on further empirical study, our coarse-grained frame-level passenger demand model has the following property.

Property 1: Stable Distribution of High Demand Slots in Frames. When all the time slots in a day are sorted according to their passenger demand, the distribution of those best n slots with highest demand in each frame is stable.



(a) Distribution of the best 240 slots with highest traffic volumes (b) Distribution of the best 720 slots with highest traffic volumes

Figure 8: The distribution of high-traffic slots ($\tau = 1$ min) in each 1 h frame when the advertising time α is 4 h or 12 h. This result is based on Dataset A.

Fig. 8 shows the distributions of the best n slots in frames of short term historical days when h and τ are 1 hour and 1 minute, respectively. Fig. 8(a) plots the distribution of the best 240 slots, i.e., 4 hour advertising time, and Fig. 8(b) plots the distribution of the best 720 slots, i.e., 12 hour advertising time. From the figures, we found that no matter n is large or small, the distribution of the best n slots in frames is stable.

According to Property 1, the proposed EveryoneCounts allocates coarse-grained frame-level AD slots into each frames based on the statistical coarse-grained demand distribution.

6 Fine-grained Entering Demand Model – RealPoisson

In this section, a non-homogeneous Poisson model with a real-time tuner, RealPoisson, is proposed to predict fine-grained passenger entering demand. The proposed model has two challenging designs to improve the prediction performance, i.e., the time-varying rate parameter based on historical traffic and an entering volume tuner based on real-time traffic.

6.1 Non-homogeneous Poisson Model

The adopted non-homogeneous Poisson model, instead of a classic Poisson model with a fixed rate parameter, supports the time-varying entering rate. Researchers used to apply a classic Poisson model to traffic volume prediction when passengers arrive at a stable rate during the prediction

duration. However, the hourly passenger entering volume greatly varies as shown in Fig. 4(a), i.e., passenger arrival rate is not stable from hour to hour. To remove the fixed-rate limitation of a classic Poisson model, we propose the non-homogeneous Poisson model, which explores historical entering traffic to achieve the time-varying rate parameter.

A non-homogeneous Poisson model is defined as a counting process $\{N(t) : t \geq 0\}$ with a rate or intensity function $\lambda(t)$. Specifically, for a short time interval of $[s, t]$, $0 < s < t$, $N(t) - N(s)$ satisfies a Poisson distribution with mean $m(t) - m(s) = \int_s^t \lambda(u)du$. In the entering demand modeling, find-grained historical average value of traffic volumes is used to construct the intensity function. The construction method first divides 24 hours in a day into slight time slices with the same length, e.g., one minute. Arriving rate in a single slice is regarded as stable and thus derived from historical average value. To get an accurate intensity function for different kinds of days, we further classify different days into weekday, weekend, and holiday. In summary, we compute the intensity function for each specific station in each specific day. For instance, the entering demand for a short time interval $[t, t + \Delta t)$ is described by a Poisson distribution with parameter $\int_t^{\Delta t} \overline{\lambda_d}(u)du$, where $\overline{\lambda_d}(t)$ is the intensity function.

6.2 Real-time Tuner

There are two reasons for introducing the real-time traffic based tuner into the entering demand model. On the one hand, there exists significant randomness in the fine-grained entering traffic, as shown in Fig. 4(b), even though the entering rates for neighboring fine-grained time slots are similar. On the other hand, the rate parameters, which are computed as the averaged historical traffic, eliminate the randomness of the entering traffic volume for each fine-grained time slot. To restore the randomness and improve the predicting accuracy of the fine-grained entering demand, we employ the real-time tuner to make use of real-time traffic information.

The real-time tuner is a classifier, predicting real-time trends (i.e., upward, downward and almost stable) of the fine-grained entering traffic volumes derived from the basic non-homogeneous Poisson model. The real-time tuner restores the randomness of the fine-grained traffic in each time slot by exploiting the impact of the trend sequence in several near-historical slots on the trend in the current slot. The tuner is trained off-line based on logistic regression using traffic trend sequences in continuous fine-grained slots. When the on-line entering demand prediction of the current slot is performed, the predicted result from the basic non-homogeneous Poission model is tuned by the tuner with the past several trends. If the predicted current trend is upward (downward), the primary prediction will be increasingly (decreasingly) tuned in proportion to the traffic volume in the last time slot. The length of trend sequences used for tuner training is set as 6 since it offers good classification accuracy by our evaluation.

Fig. 9 plots the prediction performance of the fine-grained entering demand by historical averaging (Historical), non-homogeneous Poisson prediction (PrimaryPoisson), non-homogeneous Poisson with real-time tuner (RealPoisson), and ground truth (Truth) from 08:30am to 09:00am in a weekday. It is obvious the erratic fluctuation of Truth (grey bars) reflects its randomness while the line of Historical (the dotted line in blue) shows a much more stable prediction. The PrimaryPoisson line (the dotted line in red) is neither smooth as Historical line nor fluctuates dramatically as the Truth line. RealPoisson outperforms the other two methods, especially for slots which change greatly compared with their previous slots (e.g., time 12 and time 14 circled by black ellipses). This is because our real-time tuner predicts these great changes and adjusts the predicted traffic volumes.

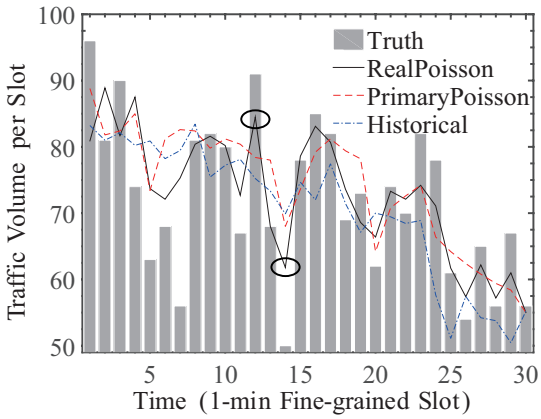


Figure 9: Performance comparison of fine-grained entering demand predictions from 08:30 am to 09:00 am in a weekday.

7 Fine-grained Exiting Demand Model – EveryoneCounts

We propose a bayesian model, EveryoneCounts, for fine-grained exiting demand modeling, based on both historical and real-time AFC records. EveryoneCounts aims to infer passenger exiting stations and time based on the particular mobility pattern for each entering passenger recorded by the AFC system. Since the passenger exiting demand in a future slot is the number of passengers arriving at the target station from other metro stations, their AFC records entering the metro system have already been recorded. We use such logged AFC records, including both the origin station and the entering time, as a condition to predict passenger arrivals, thus to obtain passenger exiting demand at slot levels. In detail, there are three steps, namely, destination prediction, travel duration prediction, and traffic volume aggregation.

7.1 Destination Prediction

For a passenger k entering the metro through station θ_k at the time μ_k^θ, we predict the destination d_k based on the recorded AFC entering information according to the predictable passenger destinations.

Property 2: Low-entropy Destination. The entropy of destination, $H(d_k)$, for a metro passenger k is low and the conditional entropy $H(d_k|\theta_k, \mu_k^\theta)$ is even much lower when the origin θ_k and starting time μ_k^θ are given.

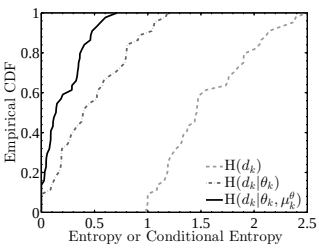


Figure 10: The CDFs of the entropy of destinations and the conditional entropy of the destinations given origins and the hour of trip starting time. The result is based on Dataset B.

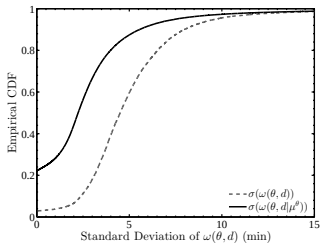


Figure 11: The CDFs of the standard deviation of travel durations given a pair of origin and destination and the hour of trip starting time. This result is based on Dataset B.

Fig. 10 plots the CDFs of $H(d_k)$, $H(d_k|\theta_k)$, and $H(d_k|\theta_k,\mu_k^\theta)$ for all the metro records of passenger k during 3 months, which are computed as follows,

$$\begin{aligned} H(d_k) &= -\sum_{d_k\in\Psi} p(d_k)\log p(d_k),\\ H(d_k|\theta_k) &= \sum_{d_k,\theta_k\in\Psi} p(\theta_k,d_k)\log\tfrac{p(\theta_k)}{p(\theta_k,d_k)},\\ H(d_k|\theta_k,\mu_k^\theta) &= \sum_{d_k,\theta_k\in\Psi,\mu_k^\theta\in\chi} p(\theta_k,d_k,\mu_k^\theta)\log\tfrac{p(\theta_k,\mu_k^\theta)}{p(\theta_k,d_k,\mu_k^\theta)}, \end{aligned} \tag{5}$$

where Ψ, the set of all the metro stations, is the support of θ_k and d_k, which can be considered as random variables, and χ = {[00:00:00, 01:00:00), [01:00:00, 02:00:00), $\cdots$, [23:00:00, 24:00:00)}, the set of hours in a day, is the support of the random variable μ_k^θ. We find from the figure that $H(d_k|\theta_k,\mu_k^\theta)$ is lower than 0.7, which means there are only $2^{0.7}$ possible destinations, compared to totally 118 stations in Shenzhen metro, for each passenger when θ_k and the hour of μ_k^θ are given. By exploring the travel history of each passenger, the destination of a future metro trip can be exactly predicted given the recorded AFC information of the origin and the starting time of the trip, i.e., finding the destinations mostly associated with the origin and the starting time in the historical data.

7.2 Travel Duration Prediction

We then predict the arrival time ν_k^d of passenger k at the predicted destination d_k based on the recorded θ_k and μ_k^θ, according to the following property of passenger mobility.

Property 3: Stable Travel Duration. The travel durations between a given pair of origin and destination, denoted by $\omega(\theta,d)$, is stable. Moreover, if the hour of the starting time μ^θ is given, the standard deviation of the travel durations for a pair of origin and destination, denoted by $\sigma(\omega(\theta,d))$, will be further reduced.

Fig. 11 plots the CDFs of $\sigma(\omega(\theta,d))$, and $\sigma(\omega(\theta,d|\mu^\theta)), \theta,d \in \Psi$, where $\omega(\theta,d|\mu^\theta)$ represents the travel durations for a given pair of stations and starting during a given hour in χ. We find from the figure that 70% of $\sigma(\omega(\theta,d|\mu^\theta))$ are lower than 3 min.

In this work, we use 90th percentile of the standard deviation to obtain an arriving time interval, instead of a fixed value, for an uncertain demand model.

7.3 Traffic Volume Aggregation

Based on the arriving time intervals of all passengers who are already in the metro network, we obtain a lower bound, i.e., $\underline{\beta^0}_{j\cdot d}$, and a upper bound, i.e., $\overline{\beta^0}_{j\cdot d}$ for exiting demand at station ψ_d during slot t_j based on all the AFC records occurring before the current slot t_{j-1}. These lower and upper bounds are used in our scheduling with robust receding horizon control as follows.

8 Scheduling by Receding Horizon Control

Based on our two-level demand model, our scheduling also has two parts, i.e., offline advertising time allocation at frame level, and online advertising time allocation at slot level.

8.1 Offline Advertising Time Allocation at Frame Level

Given n AD slots and the historical coarse-grained traffic distribution of $\frac{T}{h}$ frames, we first allocate n to the frames. The number of AD slots for frame $f_l, 1\le l\le\frac{T}{h}$, is denoted by η_l. The allocation is inspired by the Property 1 of our coarse-grained demand model, i.e., stable distribution of high demand slots in frames. So we determine the allocation $\{\eta_l\}_{l=1}^{\frac{T}{h}}$ according to the stable distribution of high-traffic slots in frames. Given the fixed n and the historical percentage, denoted by π_l, of the high-traffic slots in frame f_l, η_l is computed as the product of n and the stable historical percentage π_l, i.e., $\eta_l = n\times\pi_l$, where $\sum_{l=1}^{\frac{T}{h}}\eta_l = n$, and $\sum_{l=1}^{\frac{T}{h}}\pi_l = 1$.

8.2 Online Advertising Time Allocation at Slot Level

With the allocation of the number of the advertising slots in each frame and the predicted real-time traffic volumes in future slots in each frame, the schedule $\{\lambda_j^l\}$ are made with receding horizons of real-time AFC records, where $1\le l\le\frac{T}{h}, 1\le j\le\frac{h}{\tau}$. The pseudo code in Alg. 1 explains the main process of the scheduling.

Algorithm 1: Online AD Slot Allocation

Input $\{\eta_l\}$: the allocation of n advertising slots for each frame f_l;
Real-time card swiping records
Output $\{\lambda_j^l\}$: the schedule of all time slots in all frames
1: $l=1, j=0$
2: **for** $l=1, l<=\frac{T}{h}$(i.e., all the frames) **do**
3: **for** $j=0, j<\frac{h}{\tau}$ **do**
4: Predict the lower and upper bounds for $\{\beta_{j+1}^l,\beta_{j+2}^l,\cdots,\beta_{\frac{h}{\tau}}^l\}$ based on the updated AFC records occurring before t_j^l.
5: Solve robust advertising efficiency optimization problem proposed in Eq.(4).
6: When the current time moves into t_{j+1}^l, display advertisements if λ_{j+1}^l is 1.
7: **Continue**
8: **end for**
9: **end for**

During a frame f_l, before each slot t_j^l, the advertising schedule λ_j^l, is made by first predicting lower and upper bounds of passenger demand in the future slots in f_l, $\{\beta_{j+1}^l, \cdots, \beta_{\frac{h}{\tau}}^l\}$ with updated AFC records. Then, we solve the robust advertising efficiency optimization problem proposed in Eq.(4) by a numerical method to obtain the schedule for the rest of slots. Then, we choose to display the AD or not based on the obtained schedule.

9 Performance Evaluation

Based on one month data from Dataset B in Fig. 3, we evaluate the performance of the coarse-grained demand prediction, the fine-grained entering and exiting demand prediction, and the resulted advertising efficiency against different settings of the slot length (default: 3 min), frame length (default: 1 hour), and the length of advertising time (default: 5 hours).

We respectively compare the performance of the proposed fine-grained entering demand model(RealPoisson) and the fine-grained exiting demand model (EveryoneCounts) with the performance of the optimal scheduling and two existing approaches as follows.

- **The optimal approach (Optimal):** This approach makes the optimal schedule for each time slot to achieve the optimal advertising efficiency under the given advertising time and slot length, based on the ground truth of fine-grained traffic volumes.
- **Coarse-grained scheduling approach (Coarse-grained):** This approach selects the best coarse-grained frames with highest averaged historical traffic volumes in near historical days to display advertisements. The frames are sorted and selected to display advertisements in decreasing order of their historical frame traffic volume.
- **Historical traffic volume based approach (Historical):** This approach applies fine-grained advertising based on pure historical traffic volume instead of both historical and real-time information. Fine-grained slots are sorted and selected as advertising slots in decreasing order of the historical averaged slot traffic volume in near history.

9.1 Performance of Coarse-grained Demand Prediction

We first explore the coarse-grained frame-level AD slots allocation performance under different settings of the slot length τ and the frame length h, given the default value of α.

Fig. 12 plots the allocation error δ_η when τ varies from 1 min to 6 min and h varies from 0.5 hour to 3 hour. The allocation error δ_η is computed as the average prediction error of the number of high-traffic slots in each frame,

$$\delta_\eta = \frac{h}{T} \sum_{l=1}^{\frac{T}{h}} \frac{|\eta_l - \hat{\eta}_l|}{\max\{\eta_l, \hat{\eta}_l\}}, \tag{6}$$

where $\{\hat{\eta}_l\}$ are the true distribution of the high-traffic slots in the frames of the target day.

We can find from the figure that the allocation error of most settings are lower than 10%, especially when the slot length is 3 min and the frame length is longer, e.g., 3 h.

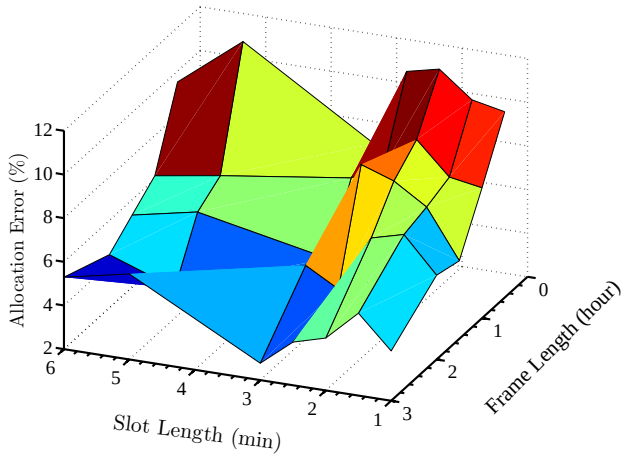


Figure 12: The allocation error of the given advertising time (five hours) into coarse-grained frames when both the length of slots and the length of frames vary.

The reason for the largely decreasing allocation error as the frame length increases is that the more coarse-grained traffic volume is more predictable. This phenomenon corresponds to the Observation 1.1 and 2.1 that coarse-grained entering and exiting demands are predictable.

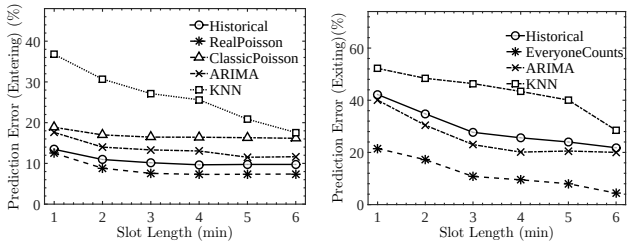
We can also find that when the slot length is 3 min, the allocation error is the lowest among those with the same frame length. The key reason is as follows. The train interval in the metro system in Shenzhen city is 3 min in rush hours and 6 min in nonrush hours. Then, the 3 min time slots naturally grouping the passengers getting off from the same train. As a result, the traffic volumes in every 3 min slot are more regular than in other length of slots. Inspired by this conclusion, we suggest the setting of slot length τ according to the train interval of the target metro station.

9.2 Performance of Fine-grained Demand Prediction

We then respectively evaluate the online fine-grained demand prediction performance of RealPoisson and EveryoneCountes. We compare the prediction error of RealPoisson with that of Historical, ClassicPoisson (homogeneous Poisson model with a fixed rate parameter), K nearest neighbors regression model (denoted by KNN in result figures), and autoregressive integrated moving average model (denoted by ARIMA in result figures) when slot length varies. We compare the performance of EveryoneCountes with that of Historical, KNN, and ARIMA under different settings of the slot length. Parameter K is set as 10 for KNN regression, and $(p, d, q) = (2, 1, 2)$ for ARIMA model, where an initial differencing step is applied to eliminate the non-stationarity. The results are plotted in Fig. 13(a) and Fig. 13(b), respectively.

From Fig. 13(a), we can find that prediction errors go down when the time slot is longer. Only when the slot is extremely short, i.e., 1 min, the prediction errors of all the models significantly increase, which accords with Observation 1.2. But the performance for 2-min and longer slots are relatively closer, which agrees with Fig. 6(a). This is because the entering demand for a typical commute metro station is relatively more regular and predictable than the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2017.2725913, IEEE Transactions on Big Data



(a) The prediction error of entering demand models.

(b) The prediction error of exiting demand models.

Figure 13: The prediction error of fine-grained demand models.

exiting demand. We can also find that the prediction errors of Historical and RealPoisson are much smaller than that of ClassicPoisson. This is because a classic Poisson model ignores the different entering rate for different period (e.g., peak and off-peak hours) in a day.

We can find from Fig. 13(b) that the prediction errors of both approaches decrease as the slot length increases and the prediction error of EveryoneCounts is on average 61.49% lower than that of Historical. When the slot length is as short as 1 min, the prediction performance of both approaches is poor while the reasons of their poor performance are different. The reason for the high prediction error of EveryoneCounts with 1 min slot length is the deviation of the individual travel duration from the average historical duration of each passenger (as shown in Fig. 11) is much larger than the 1 min slot length. While the reason for the extremely high prediction error of Historical is the uncertainty of the fine-grained traffic volume, our novel observations made in Section 3.3.2.

To show the prediction performance of typical time series analysis approaches and machine learning approaches, we evaluates the prediction error of ARIMA and KNN regression. From the figures, we can find that KNN achieves highest prediction error, which is resulted from the challenging feature selection. In the evaluations, the features to predict the passenger volume in a future slot are set as the passenger volumes in previous slots which are shown by our Observation 1.2 to be unpredictable and fluctuated. As a result, KNN performs worst. ARIMA has a similar prediction error with Historical. As expected, for the entering passenger demand, which is relatively more regular than the exiting demand, the prediction error of ARIMA is a little higher than that of Historical and for the exiting demand prediction, ARIMA performs worse than Historical.

9.3 Performance of Advertising Efficiency for Entering Passengers

The performance of the advertising efficiency for entering passengers of RealPoisson and all the compared approaches are evaluated against the impact of advertising time α, slot length τ and frame length h. The default length of a frame used in the subsection is 2 hours.

9.3.1 Impact of the Advertising Time

Fig. 14(a) plots results of advertising efficiency when the advertising time increases from 1 h to 8 h. From the picture, we can see that the proposed RealPoisson outperforms all other approaches except Optimal. The advertising efficiency decreases for all approaches as the advertising time increases. The downward tendency is due to unevenly distribution of traffic volumes at different hours in the daytime as shown in the coarse-grained entering demand of Fig. 4(a). When the advertising time is short, we can choose those hours with heavy traffic. The Coarse-grained approach has the worst performance sine it can only select complete frames or not and thus miss valuable slots with high traffic volumes in others frames. We will further discuss this issue when considering the impact of frame length.

9.3.2 Impact of the Slot Length

We then evaluate the impact of slot length on advertising efficiency with a fixed length of advertising time and frame length. The results are plotted in Fig. 14(b), where the Optimal approach has a slight decreasing while the RealPoisson and the Historical approach has a rather smaller increase when the slot length increases. It is expected that Optimal achieves its best performance when slot length is the shortest, i.e., 1 min. When the slot length is 1 min and 6 min, RealPoisson has an advertising efficiency 91.4% and 95.1% of that of the Optimal, respectively. The Coarse-grained approach has a fixed advertising efficiency since its allocation scheme only considers frame-level traffic volumes. In summary, the slot length has little impact on the advertising efficiency of all the compared approaches, resulted by the relatively regular and predictable entering demand in a typical commute station.

9.3.3 Impact of the Frame Length

We present the impact of the frame length on advertising efficiency when the frame length ranges from 0.5 h to 3 h in Fig. 14(c). The figure shows that the increasing frame length diminishes the advertising efficiency of both RealPoisson and Coarse-grained while obviously has no impact on Optimal and Historical. Coarse-grained has the fastest decrease when the frame length is enlarged since its coarse-grained selection results in larger amount of low-traffic slots in the selected frames. RealPoisson has a much slower decrease than Coarse-grained, which is benefited from the fine-grained selection avoiding those low-traffic slots in each frame. The decreasing efficiency of RealPoisson is resulted by the lack of real-time information in the traffic prediction for the future slots in a longer frame.

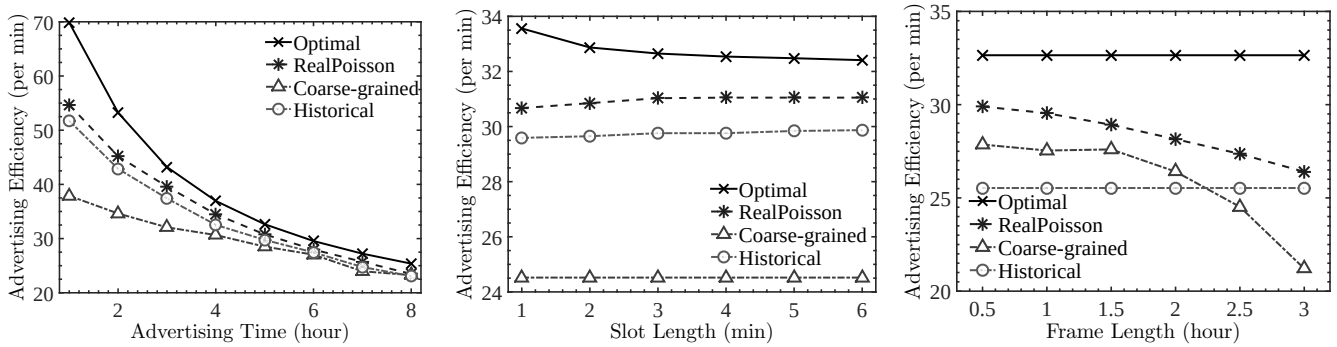
9.4 Performance of Advertising Efficiency for Exiting Passengers

The performance of the advertising efficiency for exiting passengers of EveryoneCounts and all the compared approaches are evaluated against the impact of advertising time α, slot length τ and frame length h.

9.4.1 Impact of the Advertising Time

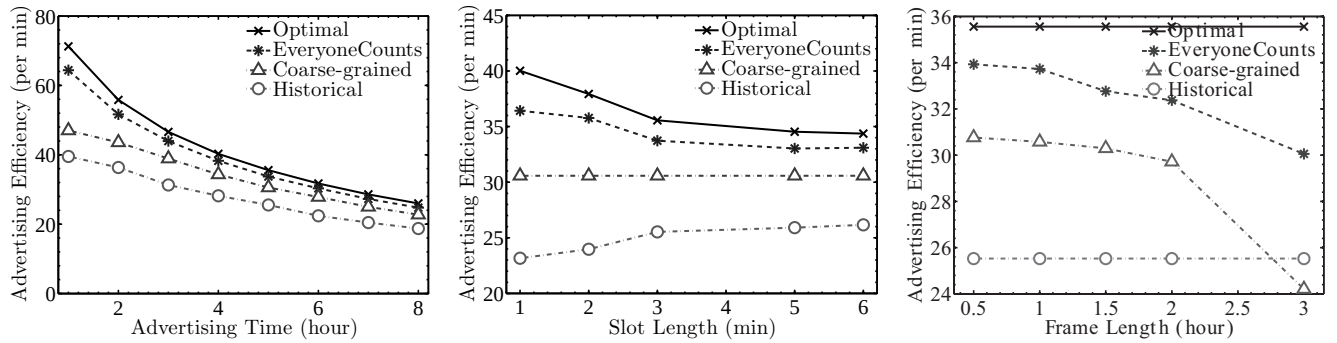
Fig. 15 and Fig. 17(a) plot the total traffic volume of the exiting passengers accumulated during all the advertising time and the advertising efficiency when the given advertising time varies from 1 h to 8 h.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2017.2725913, IEEE Transactions on Big Data



(a) The advertising efficiency when the advertising time varies from 1 h to 8 h. (b) The advertising efficiency when the slot length varies from 1 min to 6 min. (c) The advertising efficiency when the frame length varies from 0.5 h to 3 h.

Figure 14: The performance of advertising efficiency for entering passengers.



(a) The advertising efficiency when the advertising time varies from 1 h to 8 h. (b) The advertising efficiency when the slot length varies from 1 min to 6 min. (c) The advertising efficiency when the frame length varies from 0.5 h to 3 h.

Figure 17: The performance of advertising efficiency for exiting passengers.

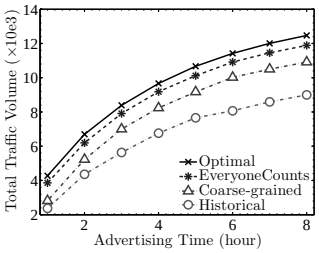


Figure 15: The total traffic volumes of exiting passengers when the advertising time varies from 1 h to 8 h.

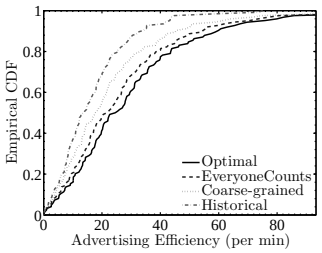


Figure 16: The empirical CDFs of the advertising efficiency in all the 118 metro stations of all the approaches.

From the two figures, EveryoneCounts performs the best among all the non-optimal approaches. As the advertising time increases, as expected, the total traffic volume increases, while the advertising efficiency decreases for all the approaches. Since the exiting traffic is not evenly distributed in the target day and those time slots with higher traffic volumes will be preferentially selected, the longer the advertising time, the lower the advertising efficiency. When the advertising time is 1 h, the total traffic volume (advertising efficiency) of EveryoneCounts is 37% and 62.9% higher than that of Coarse-grained and that of Historical, respectively. The reason for the lower advertising efficiency of Coarse-grained than that of EveryoneCounts is the fluctuated fine-grained exiting traffic volume even during rush hours, which has been introduced as Observation 2.2. Historical has the worst performance of the advertising efficiency because of the high prediction error of the fine-grained traffic volume, which is resulted by the significant temporal irregularity of fine-grained exiting traffic, even much more unpredictable than the entering traffic.

9.4.2 *Impact of the Slot Length*

Given the fixed length of the advertising time, the length of the slots has different impacts on the advertising efficiency of different approaches. Fig. 17(b) plots the advertising efficiency of all the approaches when the slot length varies from 1 min to 6 min.

Optimal and our approach have decreasing advertising efficiency while Historical has increasing advertising efficiency as the slot length increases. For Coarse-grained, since it applies the coarse-grained scheduling, the slot length has no impact on its performance. The decreasing advertising efficiency of Optimal and EveryoneCounts is resulted by the more coarse-grained advertising with longer slot length. Although EveryoneCounts has a decreasing advertising efficiency, the distance between its efficiency and that of Optimal decreases. For example, when the slot length is 1 min and 6 min, the efficiency of EveryoneCounts is 91% and 96.3% of that of Optimal, respectively. As expected, the lowest advertising efficiency of Historical when the slot length is shortest is resulted from the temporal uncertainty of fine-grained exiting traffic.

9.4.3 Impact of the Frame Length

Fig. 17(c) presents the advertising efficiency of approaches when the frame length varies from 0.5 h to 3 h. Optimal and Historical are not affected by the frame length because they both apply only fine-grained advertising. Our approach and Coarse-grained have decreasing performance with larger frame length because they both take advantage of the coarse-grained traffic certainty. However, longer frame length leads to higher traffic volume prediction error of EveryoneCounts since more and farther future exiting traffic need to be predicted in longer frames while real-time entering information for those exiting traffic is unavailable.

9.5 Global Performance

We show the advertising efficiency for exiting passengers achieved by all the approaches in all the metro stations under the default setting in Fig. 16. From the figure, we can find that EveryoneCounts has the closet performance with Optimal. The average advertising efficiency of EveryoneCounts is 26.84, which is 89.7% of the efficiency of Optimal, and 18.6% and 58.5% higher than the efficiency of Coarse-grained and Historical, respectively.

10 Discussions

In this section, we discuss some practical issues regarding to the real-world deployment of the proposed approach.

Advertisement Fairness: The proposed approach is to assist a metro system optimize its advertising efficiency given the total advertising time rented by all advertisers. Since an advertiser is charged according to the length of its advertising time, it is also important to maintain the fairness among all advertisers. Our approach can be easily adapted to take the fairness into account. When a time slot is selected as an advertising slot, each advertisement will be displayed with a probability proportional to the its length of rented time. As a result, when the rented time of an advertisement is longer, its display probability would be higher.

Passengers Contributing to Revenues: In this paper, we envision that all passengers passing by a digital screen would see advertisements. Admittedly, in reality, only a portion of passengers would see advertisements, and an even smaller portion of passengers would actually contribute to advertisers' revenues. However, these real-world factors are extremely difficult to quantify, and are out of the technical scope of this paper. Thus, we focus on the total passing by passenger traffic volume in this work.

Passengers with Temporary Cards: In a real-world metro system, there are passengers traveling with temporary cards, e.g., foreign visitors. Thus, the mobility patterns of these passengers cannot be inferred from the historical travel records. In this work, for these passengers, the destinations and the travel durations are predicted based on statistical general regularities of the passenger mobility within the metro system.

Metro Station Exits Without AFC Machines: Some metro stations might not have AFC machines at the exits, so the destinations of passengers are unavailable. Our approach is still applicable for such stations, because the destination and travel duration of a passenger trip can be estimated by exploring the entering record of the passenger's next trip [7].

Demand Modeling using Other Infrastructures: Other infrastructures, e.g., wifi routers or infrared sensors, can also be used to model passenger demand. But they typically introduce additional costs. In contrast, our approach based on AFC data did not introduce any additional costs, since AFC data are collected automatically for billing purposes.

11 Related Work

Our approach predicts the passenger traffic volume by utilizing the real-world card-swiping information recorded by the metro automatic fare collection (AFC) system. The most related topics with this work include advertising efficiency improvement, traffic volume prediction, and traffic arrival time prediction. Although smart card data have been used before [4], [5], our work is the first one on the advertising efficiency improvement based on an uncertain demand model. To the best of our knowledge, there is few, if any, research studies on the topic, so we summarize our related work from two aspects of traffic arrival time prediction and passenger demand estimation.

11.1 Arrival Time Prediction

In our design, we predict the passenger arrivals in urban metro networks, which is determined by the arrivals of metro trains. Similarly, several existing studies propose wise designs to predict the travel time of other kinds of transportation, e.g., bus. We classify such work into three categories according to the different information they use, i.e., (i) roadside sensors [8], (ii) the taxi GPS information [9], [10], [11], and (iii) the cell phone data [12], [13], [14].

Several pieces of work have been proposed to use road-size sensors, e.g., loop detectors, to predict travel time. Based on the data collected by the roadside sensors, such approaches predict the travel time by estimating speeds of vehicles. Given the estimated vehicle speed as well as the fixed length of the road segments, these studies further predict the travel time. Taghvaeeyan et al. [8] design a portable roadside sensor system which is placed next to the roadway to measure traffic. Based on longitudinally spaced sensors, the system can measure speed information and estimate travel time.

Recently, more and more researchers concentrate on the large-scale taxi GPS data. Balan et al. [9] propose a real-time trip information system that provides passengers with the expected fare and travel time. Wang et al. [10] create a citywide model for estimating the travel time of any path in real time, based on the map information and the GPS trajectories of vehicles received in current time slots as well as the history records. Westgate et al. [11] propose a regression method for travel time distribution estimation of ambulance between any two locations based on sparse GPS data.

Zhou et al. [12] employ a novel idea to predict the bus arrival time using the cell phone data. They present a bus arrival time prediction system based on the participatory sensing data provided by cell phones of bus passengers. Another work VTrack [13] also uses sensing data from

phones, e.g., the WiFi-based positioning samples, to predict traffic delay. Derrmann et al. [14] show that dwell time distributions can serve as a good predictor for travel time estimation.

11.2 Passenger Demand Estimation

The existing work on the real-time passenger demand, i.e., traffic volume estimation, mainly focuses on the volume estimation of vehicles, which can be classified into two categories, i.e., parametric and non-parametric methods. Parametric methods typically used models include local regression model [15], and Markov chain model [16], etc. Non-parametric methods include non-parametric regression [17], Bayesian networks [18] and neural networks [19].

Based on the source of the data used for prediction, the related work can also be divided into two categories, i.e., (i) the approach using roadside sensors [13], [20], [21], and (ii) the approach using taxi GPS data [22], [23], [24], [25], [26], [27], [28], [29], [30]. Singliar et al. [20] develop a probabilistic estimation model for highway networks based on the information collected from a set of traffic sensors placed around the city. Yuan et al. [22] use both historical patterns and real-time traffic information from the GPS data of taxicabs to estimate traffic conditions. Aslam et al. [23] provide model and inference procedures which can be used to analyze traffic patterns from historical data, and to estimate current traffic status from data collected in real-time. Yuan et al. [27] propose a taxi passenger demand model for taxi drivers to quickly pick up passengers to maximum their revenue. Zhang et al. [29], [30] estimate taxicab passenger demand based on a large dataset of historical demand recorded by taxicabs' historical trips and real-time information collected by roving taxicabs.

11.3 Summary

Our work solves the advertising optimization problem based on our uncertain passenger demand model with card-swiping records collected by AFC machines in a metro system. Compared with other kinds of transportation, e.g., cars or buses, the metro system has quite different properties in both metro train traveling and passenger mobility patterns. The existing approaches for vehicle traffic prediction cannot be directly used to predict metro passenger demand.

12 Conclusion and Future Work

In this paper, we propose a novel online approach for metro digital advertising to improve advertising efficiency. Our extensive empirical study and technical efforts provide a few valuable insights on metro transit networks, which are hoped to be useful for fellow researchers on similar topics. Specifically, we find that (i) coarse-grained passenger demand is regular while fine-grained passenger demand is more dynamic; (ii) passenger mobility patterns are fairly stable, and given entering station and entering time, the exiting station can be predicted with a high accuracy; (iii) travel periods between same stations are also stable in different time; (iv) it has a high accuracy to predict passenger demand by using time and stations passengers entering the metro network as conditions.

Acknowledgements

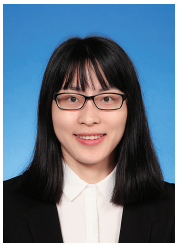
This research is supported in part by 973 Program (No. 2014CB340303), 863 Program (No. 2015AA015303), NSFC (No. 61472254, 61170238, 61472241 and 61629302), STCSM (Grant No.14511107500 and 15DZ1100305) and US NSF Grants (No. CNS-1446640 and CNS-1544887). This work is also supported by the Program for New Century Excellent Talents in University of China, the Program for Changjiang Young Scholars in University of China, the Program for Shanghai Top Young Talents, and Rutgers Research Council Award RC-18-AA-00187.

References

[1] *Metro adds 28 new digital ad displays to increase revenue*, https://beta.wmata.com/about/news/new-digital-screens.cfm.
[2] *Surge in Ridership Pushes New York Subway to Limit*, http://www.nytimes.com/2016/05/04/nyregion/surge-in-ridership-pushes-new-york-subway-to-limit.html.
[3] *New Screens Bring Added Revenue to the MTA*, http://new.mta.info/news/2014/03/13/new-screens-bring-added-revenue-mta.
[4] N. Lathia and L. Capra, "How smart is your smartcard?: Measuring travel behaviours, perceptions, and incentives," in *UbiComp*, 2011.
[5] A. A. Nunes, T. G. Dias, and J. F. e Cunha, "Passenger journey destination estimation from automated fare collection system data using spatial validation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 133–142, 2016.
[6] D. Zhang, R. Jiang, S. Wang, Y. Zhu, B. Yang, J. Cao, F. Zhang, and T. He, "Everyonecounts: Data-driven digital advertising with uncertain demand model in metro networks," in *2015 IEEE International Conference on Big Data (Big Data)*, Oct 2015, pp. 898–907.
[7] M. Trepanier, N. Tranchant, and R. Chapleau, "Individual trip destination estimation in a transit smart card automated fare collection system," *Journal of Intelligent Transportation Systems*, 2007.
[8] S. Taghvaeeyan and R. Rajamani, "Portable roadside sensors for vehicle counting, classification, and speed measurement," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 73–83, 2014.
[9] R. K. Balan, K. X. Nguyen, and L. Jiang, "Real-time trip information service for a large taxi fleet," in *ACM MobiSys*, 2011.
[10] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *ACM KDD*, 2014.
[11] B. S. Westgate, D. B. Woodard, D. S. Matteson, and S. G. Henderson, "Large-network travel time distribution estimation for ambulances," *European Journal of Operational Research*, vol. 252, no. 1, p. 322333, 2016.
[12] P. Zhou, Y. Zheng, and M. Li, "How long to wait?: Predicting bus arrival time with mobile phone based participatory sensing," in *ACM MobiSys*, 2012.
[13] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, "Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones," in *ACM SenSys*, 2009.
[14] T. Derrmann, R. Frank, S. Faye, G. Castignani, and T. Engel, "Towards privacy-neutral travel time estimation from mobile phone signalling data," in *2016 IEEE International Smart Cities Conference (ISC2)*, 2016.
[15] H. Sun, H. X. Liu, H. Xiao, R. R. He, and B. Ran, "Use of local linear regression model for short-term traffic forecasting," *Transportation Research Record: Journal of the Transportation Research Board Publisher*, 2007.
[16] G. Yu, J. Hu, C. Zhang, L. Zhuang, and J. Song, "Short-term traffic flow forecasting based on markov chain model," in *IEEE Intelligent Vehicles Symposium*, 2003.
[17] T. Zhang, L. Hu, Z. Liu, and Y. Zhang, "Nonparametric regression for the short-term traffic flow forecasting," in *Mechanic Automation and Control Engineering*, 2010.
[18] E. Castillo, J. M. Menndez, and S. Snchez-Cambronero, "Predicting traffic flow using bayesian networks," *Transportation Research Part B: Methodological*, 2008.
[19] F. Jin and S. Sun, "Neural network multitask learning for traffic flow forecasting," in *IEEE International Joint Conference on Neural Networks*, 2008.

[20] T. Singliar and M. Hauskrecht, "Modeling highway traffic volumes," in *ECML*, 2007.
[21] M. Lippi, M. Bertini, and P. Frasconi, "Collective traffic forecasting," in *Machine Learning and Knowledge Discovery in Databases*, 2010.
[22] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *ACM KDD*, 2011.
[23] J. Aslam, S. Lim, X. Pan, and D. Rus, "City-scale traffic estimation from a roving sensor network," in *ACM SenSys*, 2012.
[24] P. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi gps traces," in *Pervasive Computing*, 2012.
[25] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani, "An energy-efficient mobile recommender system," in *KDD*, 2010.
[26] Y. Huang and J. W. Powell, "Detecting regions of disequilibrium in taxi services under uncertainty," in *SIGSPATIAL*, 2012.
[27] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, "Where to find my next passenger," in *UbiComp*, 2011.
[28] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi-passenger demand using streaming data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1393–1402, 2013.
[29] D. Zhang, T. He, S. Lin, S. Munir, and J. A. Stankovic, "Dmodel: Online taxicab demand model from big sensor data in a roving sensor network," in *2014 IEEE International Congress on Big Data, Anchorage, AK, USA, June 27 - July 2, 2014*, 2014.
[30] ——, "Taxi-passenger-demand modeling based on big data from a roving sensor network," *IEEE Transactions on Big Data*, vol. PP, no. 99, pp. 1–1, 2016.

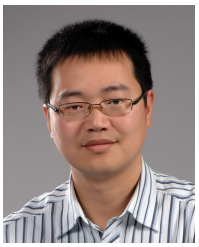
Ruobing Jiang is a PhD student in the Department of Computer Science and Engineering at Shanghai Jiao Tong University. She received her BEng. degree from the Department of Computer Science and Engineering at Southeast University in 2010. Her research interests include vehicular ad hoc networks and mobile computing. She is a student member of the IEEE.



Zhenni Feng received her B.Eng. degree in computer science from Huazhong University of Science and Technology, Wuhan, China, in 2012. She is now a Ph.D. student in the Department of Computer Science and Engineering at Shanghai Jiao Tong University. Her research interests include machine learning, spatiotemporal data mining and big data analysis.

Desheng Zhang is an assistant professor at the Department of Computer Science, Rutgers University. Previously, Desheng was offered the Senseable City Consortium Postdoctoral Fellowship from MIT, and awarded his Ph.D in Computer Science from University of Minnesota. His research includes big data analytics, mobile cyber physical systems, wireless sensor networks, and intelligent transportation systems. His research results are uniquely built upon large-scale urban data from cross-domain urban systems, including cellphone, smartcard, taxi, bus, truck, subway, bike, personal vehicle, electric vehicle, and road networks. Desheng designs and implements large-scale data-driven models and real-world services to address urban sustainability challenges.

Shuai Wang is currently a postdoc in the Department of Computer Science and Engineering at University of Minnesota. He received his Ph.D. from the Department of Computer Science and Engineering at the University of Minnesota in 2017. He received the B.S. and M.S. degrees from Huazhong University of Science and Technology, China. His research interests include Internet of Things, mobile cyber physical systems, bigdata analytics for intelligent transportation systems, and wireless networks and sensors.



Yanmin Zhu obtained his PhD from the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology in 2007. He is a Professor in the Department of Computer Science and Engineering at Shanghai Jiao Tong University. Prior to joining Shanghai Jiao Tong University, he worked as a Research Associate with the Department of Computing at the Imperial College London, UK. His research interests include vehicular networks, wireless sensor networks and mobile computing. He is a member of the IEEE and the IEEE Communication Society.

Fan Zhang is an Associate Professor of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. He received a B.S. in Communication Engineering (2002), and M.S. and Ph.D. degrees in Communication and Information System from Huazhong University of Science and Technology, Wuhan, China in 2004 and 2008, respectively. In 2008−2009 he worked as a senior research associate at City University of Hong Kong. From 2009.11 to 2011.8, he was a postdoctoral fellow at University of New Mexico and University of Nebraska-Lincoln, USA. His research topics include the cloud computing, big data, cyber-physical systems, privacy and security in cloud computing.

Tian He is currently an associate professor in the Department of Computer Science and Engineering at the University of Minnesota-Twin City. He received the Ph.D. degree under Professor JohnA.StankovicfromtheUniversityofVirginia, Virginia in 2004. Dr. He is the author and coauthor of over 150 papers in premier network journals and conferences with over 14,000 citations (H-Index 46). His publications have been selected as graduate-level course materials by over 50 universities in the United States and other countries. Dr. He has received a number of research awards in the area of networking, including five best paper awards. Dr. He is also the recipient of the NSF CAREER Award 2009 and McKnight Land-Grant Professorship. Dr. He served a few program chair positions in international conferences and on many program committees, and also currently serves as an editorial board member for six international journals including ACM Transactions on Sensor Networks. His research includes wireless sensor networks, cyber-physical systems, intelligent transportation systems, real-time embedded systems and distributed systems,supported by National Science Foundation,IBM,Microsoft and other agencies.