Optimal Rebalancing with Waiting Time Constraints for a Fleet of Connected Autonomous Taxi

Seong Ping Chuah, Shili Xiang, Huayu Wu
Institute for Infocomm Research, A*STAR
1 Fusionopolis Way, Singapore 138632
Email: {chuahsp,sxiang,huwu}@i2r.a-star.edu.sg

***Abstract*—A fleet of cooperative autonomous taxi is an emerging application of IoT in transportation industry. Unlike manned taxis that cruise on roads uncoordinated and often compete for passengers, autonomous vehicle can move cooperatively to transport passengers more efficiently. In this paper, we present a case study on an IoT application of new cooperative management technique for a fleet of autonomous taxi. In transportation network, optimal rebalancing allows sustainable flow of vehicle with a minimum number of vehicle to transport passengers flows in uneven directions. However, long waiting time to board a taxi during peak hours degrades quality of service. To tackle this issue, we extend recent advances in autonomous mobility-on-demand solution to incorporate waiting time policy. Specifically, we introduce stability and control of passenger's queues in the optimal rebalancing to confine the queues (thus waiting time in queues) to a specified range. We validate our new technique via data-driven simulations of a fleet of autonomous taxi by leveraging on Singapore's taxi dataset. Data-driven simulations demonstrate promising results of the new technique in ensuring efficient and low waiting time of taxi service for passengers.**

I. Introduction

While vehicle-to-vehicle communication systems enables vehicles to "talk" to each other, they are still manned by human drivers who make every decisions on road. Emerging technology in autonomous vehicle has significantly expanded the application of internet-of-thing (IoT) to revolutionizes the way people transit in town [1]. A fleet of internet-connected autonomous taxi can communicate and make decisions cooperatively to cater for large urban mobility needs in much more efficient way [2].

Traditional manned taxis cruise in an uncoordinated manner, and often compete among each other for passengers. Many schemes, such as [3], [4], [5], [6], [7], [8] have been proposed to help taxi drivers match with passenger demands. More recently, Xu [9] investigated social propagation effects in predicting taxi drivers' future behaviors. All these schemes, however, play only an advisory role to the taxi drivers while cooperation level among taxis remains low.

Deploying autonomous vehicles for taxi service gives rise to a new challenge in cooperative management of the autonomous vehicles on roads[10] for efficient transportation of passengers. In autonomous mobility-on-demand (AMOD) service, the autonomous vehicles coordinate with the command center on their heading directions while handling the local driving task autonomously. With optimal coordinating policy, AMOD can be more efficient in serving the mobility needs of passengers with minimal number of vehicle [10].

On the other hand, passenger's traffics often flow in uneven directions during peak hours. [11] Rebalancing routes the vehicles from the destinations to the sources of traffics to allow sustainable flow of vehicle to transport passengers. In [10], [12], AMOD service was modeled in queuing theoretical framework, where autonomous taxis are assumed to pick-up/drop-off passengers at a number of stations. Optimal rebalancing policy was then formulated to serve the passenger's flows with minimal rebalancing traffics. However, passengers' waiting times and the build-up of queues [13], [14] at stations were not explicitly captured, analyzed and constrained.

Cooperative management of autonomous taxi fleet improves the service efficiency, and is crucial to the successful deployment of this new IoT application. However, the lack of operational trials and data from true autonomous taxi fleet poses challenges in fleet management and optimization. In this paper, we leverage on dataset from existing manned taxis in Singapore to conduct data-driven simulations, and validate the new cooperative management technique for AMOD deployment. Simulations based on real-world dataset render us more confidence for future deployment of AMOD service.

The paper is organized as follows. We describe the queuing theoretical framework in Section II. In Section III, we formulate the optimal rebalancing with waiting time policy as an optimization. We present the solution procedure in IV. We showcase a Singapore's case study in Section V. The paper is concluded in Section VI

II. Model Description

We consider a fleet of autonomous taxi roaming on roads to provide transportation service to passengers. Let there be a set $\mathcal{N}$ of stations with substantial demand for taxi service. These stations can be identified as the active points of taxi pick-up/drop-off via clustering of taxi rides within the area.

Passenger demands arrive at station $i \in \mathcal{N}$ according to a Poisson arrival process with rate λ_i, and request for taxi service destined to another station $j \in \mathcal{N}$ with the probability of p_{ij} where $p_{ij} \in \mathbb{R},\ \sum_j p_{ij} = 1,\ i \neq j, p_{ii} = 0\ \forall\ i \in \mathcal{N}$. Upon arrival at a station, the passenger takes the taxi service if autonomous vehicles are available at the station. If no autonomous vehicle is available in the station, the passenger queue to wait for taxi ride in first-in-first-out manner. Upon

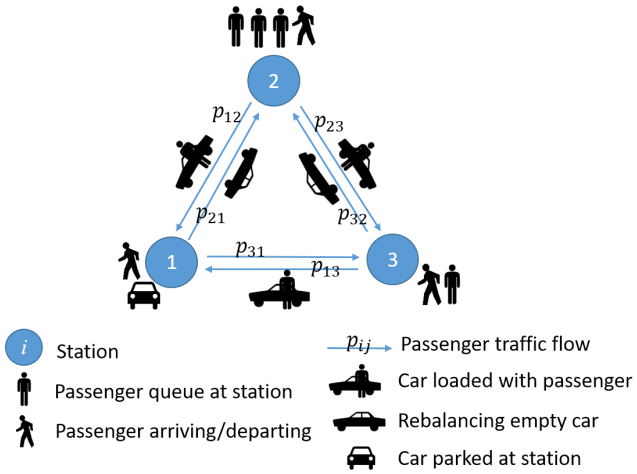


Fig. 1. Queueing network illustration for $\mathcal{N} = 3$

arrival of a vehicle, the first passenger in queue will be serviced. On the hand, if a vehicle arrives at a station without any passenger waiting in queue, the vehicle is parked at the station until an arrival of new passenger demand at the station, or instruction from central server to reroute to another station. Let d_{ij} be the shortest travel distance from station i to station j. We assume that the travel time to follow exponential distribution with mean t_{ij}. When their destinies are reached, passengers leave the queuing system. Figure 1 illustrate an example of the model description.

Few notes to highlight: First, these stations are defined solely for the purpose of network modeling, not in physical world like bus stops. Passengers demands are registered to their nearest stations for accurate modeling of traffic flows and fast taxi dispatch. Second, while [12] featured impatient passengers who leave if no vehicle is available at the station upon arrival, our model captures the true scenario where passengers and taxis often wait in (sometime, long) queues whenever there is shortage of the other side. Capturing this setup allows us to address the waiting time of passengers before getting the taxi service. Third, we assume that there is sufficient parking space for the vehicles in all stations. As we aim to deploy minimum number of vehicles, we show in the experiment that only parking space for a few vehicle is needed. Fourth, in the queuing theoretical framework, the travel times between stations is often assumed to be exponentially distributed to simplify the analysis substantially. Although travel times in practice may not follow exponential distribution and can be predicted such as [15], it has been found to have limited impact on the accuracy in practice [16], [12]. We follow the same assumption for tractable analysis. As we shall show, the steady-state mean of travel times matter more than the distribution in our rebalancing and control of queuing model. Finally, congestion routing is implicitly represented in the model by the mean traveling time t_{ij} with distance d_{ij}. Unlike a closed Jackson model [17] described in [12] where passengers move between stations without leaving or new passengers joining in, we model the network as an open cascade of queuing system, where varying flow rates of passenger between stations can be implicitly represented by varying λ_i and p_{ij}.

III. Rebalancing with Waiting Time Policy

In this section, we formulate the optimal rebalancing with waiting time policy for a fleet of autonomous vehicles.

A. Optimal Rebalancing

Rebalancing of vehicles are archived by formulating flows of 'virtual' passengers that, together with the flows of real passengers, balances the vehicles across the network. Similar to the real passengers, the virtual passengers too arrive at each station in Poisson arrival process at the rate ψ_i, and route to station j with probability q_{ij} where where $q_{ij} \in \mathbb{R},\ \sum_j q_{ij} = 1,\ i \neq j,\ q_{ii} = 0,\ \forall\ i \in \mathcal{N}$.

To balance the network, we require that the arrival and departure rates of vehicle at all stations tally, ie.

$$\underbrace{\lambda_i + \psi_i}_{\text{departure}} = \underbrace{\textstyle\sum_{j \neq i} (\lambda_j p_{ji} + \psi_j q_{ji})}_{\text{arrival}} \quad \forall\ i \in \mathcal{N} \tag{1}$$

where the departure rate of vehicle is equivalent to the sum of arrival rate of the real and virtual passenger demands. Since vehicles with virtual passengers are actually traveling without passenger (thus without revenue), we want minimal virtual passengers flow in traveling distance for cost saving. Thus, the rebalancing problem is to minimize the traveling distances of virtual passengers while satisfying (1).

$$\min_{\psi_i, q_{ij}} \quad \textstyle\sum_{ij} \psi_i q_{ij} d_{ij} \tag{2}$$

$$\text{s.t.} \quad \lambda_i + \psi_i = \textstyle\sum_{j \neq i} (\lambda_j p_{ji} + \psi_j q_{ji}) \tag{3}$$

$$\textstyle\sum_j q_{ij} = 1,\ q_{ij} \geq 0,\ \psi_i \geq 0 \tag{4}$$

Here, the rebalancing problem is formulated to minimize the traveling distance of rebalancing flow, hence d_{ij} in (2).

B. Queue Stability and Waiting Time Control

While (2)-(4) ensures the balance of network, it however does not ensure the stability of queues in the network. When a passenger demand arrives at a station without any vehicle available, the passenger waits in the queue until her turn to be served. Thus, taxi service for passengers are delayed, building up a backlog of passenger demands at all stations. Keeping queues bounded is necessary to ensure a limited waiting time for passengers.

To analyze, we model the passenger queues at stations as M/M/1 queues [17], since passenger arrivals are Poisson, and following the assumption of exponentially distributed travel times between stations. The network of queues at all stations therefore form a series cascade of M/M/1 queues. As passenger arrivals at a station are Poisson with rate $\lambda_i + \psi_i$, according to Burke's Theorem [17], the outputs of passengers from the network of queues at all stations are also Poisson with the same mean rate as inputs of passengers, which is $\sum_{j \neq i} (\lambda_j p_{ji} + \psi_j q_{ji})$. Notice that the end of taxi service

for passengers in the vehicle implies arrival of taxi service for passengers in queue at the station. Mean service rate of queue at a station is therefore $\sum_{j\neq i}(\lambda_j p_{ji} + \psi_j q_{ji})$. Here, the passengers include both real and virtual passengers. But as virtual passengers are for rebalancing, waiting time in queue is not applicable to virtual passengers. For the real passengers, the mean service rate of taxi service at station i is thus

$$\mu_i = \sum_{j\neq i}(\lambda_j p_{ji} + \psi_j q_{ji}) - \psi_i \tag{5}$$

which is the total arrivals of vehicle with real passengers and net arrivals/departure of rebalancing vehicles. Note that following the rebalancing policy (1), we have $\lambda_i = \mu_i \; \forall \; i \in \mathcal{N}$ for real passengers.

For M/M/1 queue, the steady state mean L_i of queue length l_i is given as

$$L_i = \lim_{t\to\infty} E\left[l_i(t)\right] = \frac{\gamma_i}{1-\gamma_i}, \quad \gamma_i \triangleq \frac{\lambda_i}{\mu_i} \tag{6}$$

where γ_i is the traffic intensity at station i. Obviously, a queue goes unbounded $L_i \to \infty$ as $\gamma_i \to 1$. Thus, $\lambda_i < \mu_i$ is required to keep the queue bounded. (1) does not ensure bounded queue and a limited waiting time for passengers. To solve, we let $q_{ii} \geq 0$ such that

$$\mu_i = \sum_{j\neq i}(\lambda_j p_{ji} + \psi_j q_{ji}) - \psi_i + \psi_i q_{ii} > \lambda_i \tag{7}$$

Here, $q_{ii} \geq 0$ practically means that a fraction of virtual passenger arrival ψ_i destined to the same station, and allow higher overall arrival rate of vehicle to serve the real passengers. In addition to $\psi_i q_{ij},\ j \neq i$ that rebalances the network, $\psi_i q_{ii}$ is the additional deployment of vehicle to keep the queues bounded at station i.

While maintaining a same $\gamma_i \leq 1$ for all stations ensures the same steady state mean queue length, it does not however ensure the same steady state mean waiting time for all passengers. Specifically, the probability that the waiting time t_{w} in an M/M/1 queue is no more than $t_{\max,i}$ is given by

$$P(t_{\mathrm{w},i} \leq t_{\max}) = 1 - \gamma_i e^{-\mu_i(1-\gamma_i)t_{\max}} \tag{8}$$

Notice that $P(t_{\mathrm{w},i} \leq t_{\max}) \to 0$ as $\gamma_i \to 1$ however large $t_{\max}$ is. (8) also shows that for the same level of γ_i, passengers in stations with higher μ_i have shorter waiting time in queue. We thus introduce waiting time constraints $P(t_{\mathrm{w},i} \leq t_{\max}) \geq \epsilon$ to all stations. Following (7)-(8), the constraint is expressed as

$$\begin{aligned}&\left(\lambda_i + \psi_i - \textstyle\sum_{j\neq i}(\lambda_j p_{ji} + \psi_j q_{ji}) - \psi_i q_{ii}\right) t_{\max} + \ln(\lambda_i) \\ &- \ln\left(\textstyle\sum_{j\neq i}(\lambda_j p_{ji} + \psi_j q_{ji}) + \psi_i q_{ii} - \psi_i\right) \leq \ln(1-\epsilon)\end{aligned} \tag{9}$$

Optimization formulation (2)-(4) is thus reformulated as

$$\min_{\psi_i, q_{ij}} \quad \textstyle\sum_i \left(\sum_j (\psi_i q_{ij} d_{ij}) + w\psi_i q_{ii}\right) \tag{10}$$

$$\text{s.t.} \quad \text{(3), (8)-(9), (4)} \tag{11}$$

Here, in addition to rebalancing the network, we also seek the minimum additional deployment of vehicle $\psi_i q_{ii}$ such that $P(t_{\mathrm{w},i} \leq t_{\max}) \geq \epsilon$. Note that $d_{ii} = 0$ by definition, while w is a weight in the cost function.

IV. Solution Algorithm

Optimization (10)-(11) is in general non-convex where efficient algorithm is not readily available. The waiting time constraints (9) impose nonlinear constraints to the formulation that is otherwise a linear programming problem as shown in (2)-(4). The rebalancing formulation aims to achieve $\lambda_i = \mu_i$ and (5) with minimal $\psi_i q_{ij},\ i \neq j$ flow; while queue stability control aims to meet the waiting time constraints by having $\psi_i q_{ii} + \mu_i \to \mu_i > \lambda_i$ with minimal $\psi_i q_{ii}$. (10)-(11) can be solved sequentially by balancing the network first followed by the queue control. Let $\beta_{ij} = \psi_i q_{ij}$, the rebalancing task can be solved efficiently as a linear programming problem:

Optimal Rebalancing:

$$\min_{\beta_{ij}} \quad \textstyle\sum_{i\neq j} \beta_{ij} d_{ij} \tag{12}$$

$$\text{s.t.} \quad \textstyle\sum_{j\neq i}(\beta_{ij} - \beta_{ji}) = \sum_{j\neq i} \lambda_j p_{ji} + \lambda_i \tag{13}$$

$$\beta_{ij} \geq 0 \tag{14}$$

By substituting the optimal $\beta^*_{ij},\ i \neq j$ into (8)-(9), the waiting time control task can be formulated as a convex optimization:

Waiting Time Policy:

$$\min_{\beta_{ii}} \quad \textstyle\sum_i \beta_{ii} \tag{15}$$

$$\begin{aligned}\text{s.t.} \quad & -\ln\left(\textstyle\sum_{j\neq i}\left(\lambda_j p_{ji} + \beta^*_{ji} - \beta^*_{ij}\right) + \beta_{ii}\right) \\ & -\beta_{ii} t_{\max} + \ln(\lambda_i) \leq \ln(1-\epsilon)\end{aligned} \tag{16}$$

$$\beta_{ii} \geq 0 \tag{17}$$

Furthermore, it can be seen that the minimal non-negative β^*_{ii} always lies at the equality of (16). Solving optimization (15)-(17) can be reduced to solving $|\mathcal{N}|$ independent non-linear equality (16). With the optimal solution β^*_{ij}, we can obtain $\psi^*_i = \sum_j \beta^*_{ij}$ and $q^*_{ij} = \frac{\beta^*_{ij}}{\psi^*_i}$ respectively.

V. Experiments

We validate our framework via data-driven simulations based on the real traffic scenario. We first describe preprocessing of traffic data in section V-A, followed by the real-time rebalancing and waiting time control policy in section V-B. Finally, we validate the technique in section V-C.

A. Mining of Taxi Dataset

Large scale deployment of autonomous taxi is yet to be launched. Nevertheless, existing taxi service data can accurately reflect the spatial-temporal demand pattern of passengers. Specifically, we use the dataset generated by currently manned taxis to identify all necessary parameters such as the set $\mathcal{N}$ of stations, passenger arrivals rates λ_i, transition probability between stations p_{ij}, and mean travel times between stations t_{ij} etc. The taxi dataset we use comprises roughly 7.1 millions taxi rides recorded by 15028 taxis in Singapore. The dataset contains the GPS records and the time stamps of pick-up and drop-off locations of each taxi ride. We first clean the dataset by removing entries (roughly 20 out of 7.1 million) with erroneous record, such as GPS location on sea.

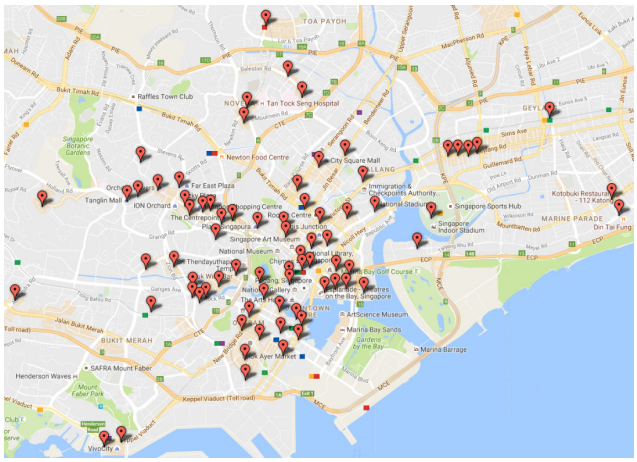


Fig. 2. The set $\mathcal{N}$ of stations identified from clustering of taxi dataset

Travel pattern usually varies according to the day of a week. Friday evening (5:30pm - 7:30pm) is among the busiest periods on Singapore's roads where long queues for taxi is not uncommon. To determine the set $\mathcal{N}$ of stations, we perform clustering on the dataset within this period of time such that every cluster represents a station with high passenger volume. As pick-up and drop-off locations are usually along roadside or buildings, the clusters are often non-convex. Thus, we use density-based method DBSCAN in the clustering. DBSCAN may return some large clusters that stretch across several junctions ranged hundreds of meters. We perform K-means clustering to break up these large clusters into smaller clusters, such that all taxi pick-up and drop-off points are within roughly 150 meters from the clusters' centroids. From the clustering, we identify the set of clusters $|\mathcal{N}| = 76$ with the highest traffic volumes. These clusters are mostly located in downtown Singapore, as shown in Fig. 2 Based on these 76 clusters, we obtain necessary parameters such as passenger arrivals rates λ_i at each station, transition probability between stations p_{ij}, and mean travel times between stations t_{ij}. We obtain the distance d_{ij} of the best routes between stations from Singapore Journey Planner [18] developed at Institute for Infocomm Research Singapore.

B. Real-Time Rebalancing and Waiting Time Policy

In section III and IV, we formulate and solve the optimal rebalancing with waiting time policy in long term steady state analysis. In practice, rebalancing and control policy can be dynamically formulated based on the real-time information of the state of arrival and queue at each station. Specifically, at station i of time t, let $Q_{\mathrm{v},i}(t)$ be the number of vehicles available, including vehicles parked at the station and the vehicles heading towards the station to drop-off passengers; while $Q_{\mathrm{c},i}(t)$ be the number of passenger queued at the station. Let $B_{ij}(t), i \neq j$ be the integer number of vehicle to rebalance from station i to station j. To rebalance and to control queue, depending on the total arrivals of passenger and the available taxi on the network at time t, the constraints are written as

$$Q_{\mathrm{v},i}(t) + \textstyle\sum_j B_{ji}(t) \geq Q_{\mathrm{c},i}(t) + \mathrm{round}(\beta^*_{ii}) + \sum_j B_{ij}(t) \quad \text{if} \quad \sum_i Q_{\mathrm{v},i}(t) \geq \sum_i \left(Q_{\mathrm{c},i}(t) + \mathrm{round}(\beta^*_{ii})\right) \tag{18}$$

$$Q_{\mathrm{v},i}(t) + \textstyle\sum_j B_{ji}(t) \leq Q_{\mathrm{c},i}(t) + \mathrm{round}(\beta^*_{ii}) + \sum_j B_{ij}(t) \quad \text{if} \quad \sum_i Q_{\mathrm{v},i}(t) \leq \sum_i \left(Q_{\mathrm{c},i}(t) + \mathrm{round}(\beta^*_{ii})\right) \tag{19}$$

(18) is the rebalancing condition where vehicles available in queues exceed the passengers waiting in queues; and vice versa for (19). Here, a rounded integer number of β^*_{ii} is allocated for each unit of time to ensure stability of queues ($\mu_i \geq \lambda_i$) and limited waiting time ($P(t_{\mathrm{w},i} \leq t_{\max}) \geq \epsilon$). The rebalancing and control policy can be formulated as a integer linear programming problem to be solved every interval:

$$\min_{B_{ij}(t)} \quad \textstyle\sum_{i \neq j} B_{ij}(t) d_{ij} \tag{20}$$

$$\text{s.t.} \quad \text{either } (18) \text{ or } (19) \tag{21}$$

$$B_{ij}(t) \geq 0 \ , \ \textstyle\sum_j B_{ij}(t) \leq Q_{\mathrm{v},i}(t) \tag{22}$$

Note that (20)-(22) bears some resemblance to (12)-(14). The second inequality in (22) is introduced to ensure causality, where all rebalancing vehicles should be physically at the stations at time t when they depart to their next stations.

C. Evaluation of The Rebalancing and Waiting Time Policy

We simulate a fleet of autonomous taxis providing taxi services in Singapore's downtown during Friday evening peak hours. Specifically, we assume that a fleet of autonomous taxi plying within the neighborhood of 76 stations based on the parameters obtained from the taxis dataset in section V-A. Real-time rebalancing and waiting time policy is updated once every minute based on the real-time infos at each station. Integer linear programming (20)-(22) is solved for every minute. The computation load is low as (20)-(22) is solved only once for all $B^*_{ij}(t)$ of 76 stations, while β^*_{ij} are only updated occasionally when arrival rates at stations vary. In practice, $B^*(t)$ is mostly a sparse matrix.

In simulations, we set $t_{\max} = 3$ minutes to match the service quality of public buses in Singapore which are very frequent during peak hours. We set $\epsilon = 0.95$ such that not more than 5% of the passengers would wait longer than 3 minutes to board a taxi. At $t = 0$, we distribute the vehicles proportionately among 76 stations according to

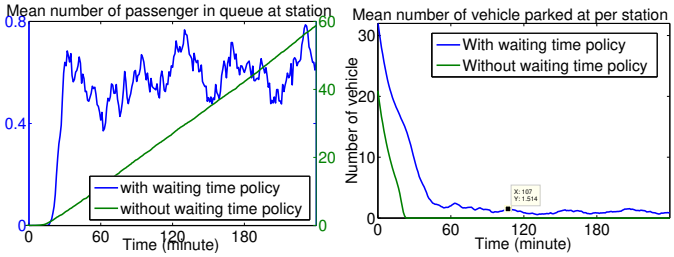
$$Q_{\mathrm{v},i}(t=0) = E[t] \cdot (\lambda_i + \beta^*_{ii}) \tag{23}$$

where $E[t]$ is the expected travel time computed as

$$E[t] = \textstyle\sum_i \left(\tilde{\lambda}_i \cdot E_i[t]\right) \quad , \quad E_i[t] = \sum_j \left(t_{ij} \cdot \tilde{p}_{ij}\right) \tag{24}$$

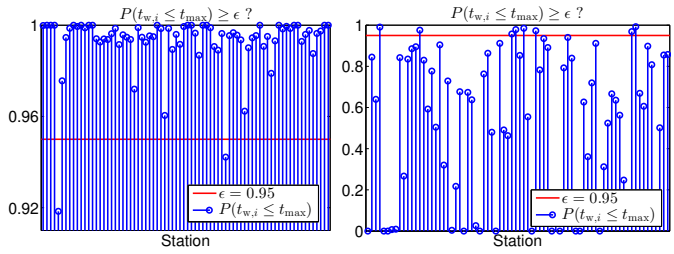
$$\tilde{\lambda}_i = \frac{\lambda_i + \psi^*_i}{\sum_i (\lambda_i + \psi^*_i)} \quad , \quad \tilde{p}_{ij} = \frac{\lambda_i p_{ij} + \psi^*_{ij} q^*_{ij}}{\sum_{ij} \left(\lambda_i p_{ij} + \psi^*_{ij} q^*_{ij}\right)} \tag{25}$$

We benchmark our rebalancing with queue control policy with that of without queue control policy in [12], where $\beta_{ii} = 0 \ \forall \ i \in \mathcal{N}$. We simulate the AMOD for an interval of 240 minutes, for 10 rounds with different random number



(a) Number of passenger queued at a station (b) Number of vehicles parked at a station

Fig. 3. Average length of queues per station in the network



(a) With waiting time policy (b) Without waiting time policy

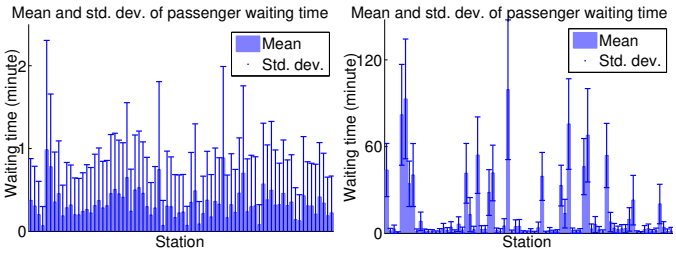
Fig. 4. The plots examine whether more than 95% of the passenger waited less than $t_{\max}$(3 minutes) to board a taxi at each station

generation seeds. The results shown below are the averages of the 10 rounds of simulations.

Fig.3(a) show the build-ups of passenger queue and vehicle queue at stations as time goes. Clearly, rebalancing without queue stability control leads to unbounded passenger queue length. Under rebalancing with queue stability control, the network reaches steady state of after roughly 1 hour where on average, less than 1 passenger is queueing at a station. Fig.3(b) shows the average queue length of vehicle at all stations. At the first hour, vehicle queue length decreases steadily as passengers arrive and take the taxi service without waiting. Without waiting time policy ($\beta_{ii} = 0$), vehicle queue lengths at all stations drop to zero as vehicles struggle to cope with increasing passengers in queues. Whereas, rebalancing policy with waiting time policy ensures that on average roughly 1.15 vehicles is available at any station at any steady state time to ensure short waiting time for passengers.

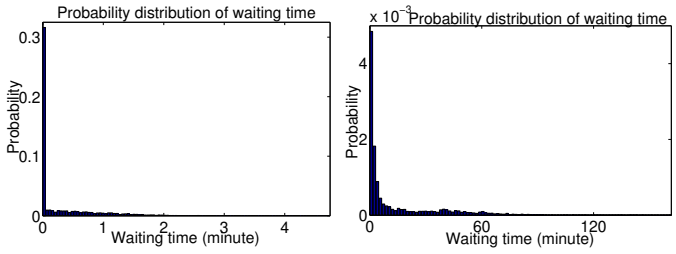
Fig.4 examines quality of service (QoS), defined as whether $P(t_{\mathrm{w},i} \le t_{\max}) \ge \epsilon$ is true at each station. Rebalancing with waiting time policy ensures that almost all stations achieve the QoS standard, except two stations whose $P(t_{\mathrm{w},i} \le t_{\max})$ is slightly lower than ϵ. Without the waiting time policy however, only a few stations achieve the QoS standard. At some stations where $P(t_{\mathrm{w},i} \le t_{\max}) \approx 0$, passengers are almost surely to wait longer than $t_{\max}$ before a taxi are available to them.

Fig.5 compares the waiting time statistics. With the waiting time policy, means of passenger's waiting times are confined within 1 minute for all stations. Whereas under rebalancing without waiting time policy, waiting time varies widely across different stations. Passengers experience long waiting times at some stations as passengers queues build up. At some stations,



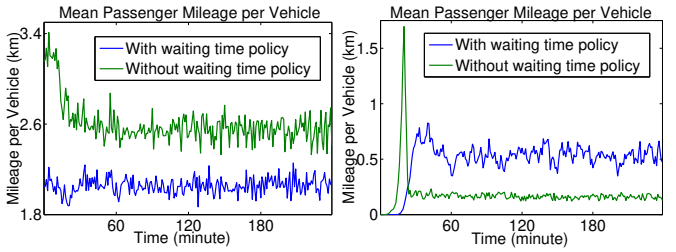
(a) With waiting time policy (b) Without waiting time policy

Fig. 5. Statistics of passengers' waiting time



(a) With waiting time policy (b) Without waiting time policy

Fig. 6. Probability distribution of passenger's waiting times



(a) Passenger mileage (b) Rebalancing mileage

Fig. 7. Mean mileages of vehicles with passengers and without (rebalancing)

passengers waited more than 30 minutes on average.

Fig.6 compares the probability distributions of waiting time of all stations. Rebalancing with waiting time policy constrains most waiting times within $t_{\max}$. Without the waiting time policy, waiting times spread across wide range from 0 seconds to more than 2 hours.

Short waiting time for passengers does not come without a cost. To cater for the similar volume of passenger demand under the selected QoS parameters ($t_{\max} = 3$ minutes, $\epsilon = 0.95$), rebalancing with waiting time policy requires 54% more vehicle than that without waiting time policy. Nevertheless, rebalancing without waiting time policy can serve less passenger volume, since inpatient passengers leaving long queues implies the loss of service opportunity. Assuming that all passengers are patient, Fig.7 shows the mean mileages per vehicle of (a) vehicles transporting passengers and (b) rebalancing vehicles. Rebalancing with waiting time policy records lower mean passenger mileage per vehicle due to its larger fleet size, but higher mean rebalancing mileage per vehicle to ensure good QoS. Mean rebalancing mileage is zeros initially as

no rebalancing is required when vehicles are available at all stations. Mean rebalancing mileage increases concurrently as passenger queues build up. Notice the slight drop of passenger mileage and the rise of rebalancing mileage around 30 minutes as passengers wait in queues for rebalancing vehicles to arrive. With the waiting time policy, the passengers mileage and rebalancing mileage are maintained consistently. Without the waiting time policy, passenger mileage drops quickly as vehicle services lag behind passenger queues. Rebalancing traffics could not catch up with the build-up of queues. Eventually, mean rebalancing mileage drops to a low level as many arriving vehicles are boarded by passengers queuing at every station.

VI. Conclusion

We present a case study on the IoT application of new cooperative management technique for autonomous taxis. Long waiting time to board a taxi degrades QoS and wastes passenger time. To tackle this issue, we experiment the optimal rebalancing framework that incorporates passenger's waiting time policy for a fleet of autonomous taxi. We demonstrate the importance of stability and control of queue for short waiting time of passengers. We model the AMOD service as a network of M/M/1 queues, and formulate a passenger waiting time constraint as part of the optimal rebalancing policy. We validate our model with realistic data-driven simulations based on the parameters extracted from dataset of manned taxis. For future work, the optimal rebalancing policy should incorporate the dynamics of congestion routing of road network and varying arrival rates of passengers in the queuing system. These factors pose difficult challenges to the rebalancing of vehicles as the passengers' waiting times are more unpredictable.

VII. Acknowledgment

This research is supported by the National Research Foundation, Prime Ministers Office, Singapore, under NRF-NSFC Joint Research Grant Call on Data Science (NRF2016NRF-NSFC001-113).

References

[1] Y. U. Devi and M. S. S. Rukmini, "Iot in connected vehicles: Challenges and issues 2014; a review," in *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, Oct 2016, pp. 1864–1867.

[2] N. Kolbe, S. Kubler, J. Robert, Y. Le Traon, and A. Zaslavsky, "Towards semantic interoperability in an open iot ecosystem for connected vehicle services," in *2017 Global Internet of Things Summit (GIoTS)*, June 2017, pp. 1–5.

[3] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-finder: A recommender system for finding passengers and vacant taxis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 10, pp. 2390–2403, Oct 2013.

[4] Meng Qu, Hengshu Zhu, Junming Liu, Guannan Liu, and Hui Xiong, "A cost-effective recommender system for taxi drivers," in *Proceedings of ACM SIGKDD*, New York, USA, 2014, pp. 45–54, ACM.

[5] Shiyou Qian, Jian Cao, Frédéric Le Mouël, Issam Sahel, and Minglu Li, "Scram: A sharing considered route assignment mechanism for fair taxi route recommendations," in *Proceedings of the 21th ACM SIGKDD*, New York, NY, USA, 2015, KDD '15, pp. 955–964, ACM.

[6] F. Miao, S. Han, S. Lin, J. A. Stankovic, D. Zhang, S. Munir, H. Huang, T. He, and G. J. Pappas, "Taxi dispatch with real-time sensing data in metropolitan areas: A receding horizon control approach," *IEEE Trans. Auto. Sc. & Eng.*, vol. 13, no. 2, pp. 463–478, April 2016.

[7] Meng-Fen Chiang, Tuan-Anh Hoang, and Ee-Peng Lim, "Where are the passengers?: A grid-based gaussian mixture model for taxi bookings," in *Proceedings of the 23rd SIGSPATIAL*, New York, NY, USA, 2015, GIS '15, pp. 32:1–32:10, ACM.

[8] S. Ma, Y. Zheng, and O. Wolfson, "Real-time city-scale taxi ridesharing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 7, pp. 1782–1795, July 2015.

[9] Tong Xu, Hengshu Zhu, Xiangyu Zhao, Qi Liu, Hao Zhong, Enhong Chen, and Hui Xiong, "Taxi driving behavior analysis in latent vehicle-to-vehicle networks: A social influence perspective," in *Proceedings of the 22nd ACM SIGKDD*, New York, USA, 2016, pp. 1285–1294, ACM.

[10] R. Zhang, K. Spieser, E. Frazzoli, and M. Pavone, "Models, algorithms, and evaluation for autonomous mobility-on-demand systems," in *2015 American Control Conference (ACC)*, July 2015, pp. 2573–2587.

[11] D. Shao, W. Wu, S. Xiang, and Y. Lu, "Estimating taxi demand-supply level using taxi trajectory data stream," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, Nov 2015, pp. 407–413.

[12] Rick Zhang and Marco Pavone, "Control of robotic mobility-on-demand systems: A queueing-theoretical perspective," *The International Journal of Robotics Research*, vol. 35, no. 1-3, pp. 186–203, 2016.

[13] Yu Lu, Shili Xiang, Wei Wu, and Huayu Wu, "A queue analytics system for taxi service using mobile crowd sensing," in *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, New York, NY, USA, 2015, UbiComp/ISWC'15 Adjunct, pp. 121–124, ACM.

[14] Yu Lu, Shili Xiang, and Wei Wu, "Taxi queue, passenger queue or no queue? - a queue detection and analysis system using taxi state transition," in *EDBT*, 2015.

[15] A. Deep Singh, W. Wu, S. Xiang, and S. Krishnaswamy, "Taxi trip time prediction using similar trips and road network data," in *2015 IEEE International Conference on Big Data (Big Data)*, Oct 2015, pp. 2892–2894.

[16] Richard C. Larson and Amedeo R. Odoni, *Urban Operations Research*, Prentice-Hall, 1981.

[17] Donald Gross, John F. Shortle, James M. Thompson, and Carl M. Harris, *Fundamentals of Queueing Theory*, Wiley-Interscience, New York, NY, USA, 4th edition, 2008.

[18] L. Yu, D. Shao, and H. Wu, "Next generation of journey planner in a smart city," in *2015 IEEE ICDM Workshop*, Nov 2015, pp. 422–429.