CityLines: Hybrid Hub-and-Spoke Urban Transit System

Guanxiong Liu
Worcester Polytechnic Institute (WPI)
gliu2@wpi.edu

Yanhua Li
Worcester Polytechnic Institute (WPI)
yli15@wpi.edu

Zhi-Li Zhang
University of Minnesota, Twin Cities
zhzhang@cs.umn.edu

Jun Luo
Machine Intelligence Lab
Lenovo Group Limited, Hong Kong
jluo1@lenovo.com

Fan Zhang
Shenzhen Inst. of Advanced Tech.
Shenzhen, China
zhangfan@siat.ac.cn

ABSTRACT

Rapid urbanization has posed significant burden on urban transportation infrastructures. In today's cities, both private and public transits have clear limitations to fulfill passengers' needs for quality of experience (QoE): Public transits operate along fixed routes with long wait time and total transit time; Private transits, such as taxis, private shuttles and ride-hailing services, provide point-to-point transits with high trip fare. In this paper, we propose *CityLines*, a transformative urban transit system, employing hybrid hub-and-spoke transit model with shared shuttles. Analogous to Airlines services, the proposed CityLines system routes urban trips among spokes through a few hubs or direct paths, with travel time as short as private transits and fare as low as public transits. CityLines allows both point-to-point connection to improve the passenger QoE, and hub-and-spoke connection to reduce the system operation cost. Our evaluation results show that CityLines framework can achieve both short travel time and high ride-sharing ratio.

CCS CONCEPTS

•**Information systems Geographic information systems;** *Computing platforms;*

KEYWORDS

Hub-and-spoke Network, urban computing, spatio-temporal data analytics

1 INTRODUCTION

Conventionally, there are two primary models of urban transport systems, namely, *public transit services* such as buses, subway, and *private passenger services* such as taxis, shared shuttles, ride-hailing services (e.g., Uber or Lyft). Both systems have limitations in fulfilling passengers' demands or "quality-of-experience" (QoE), especially during peak demand hours, due to the following fundamental trade-offs in transit service efficiency and costs. Private transits provide exclusive (non-stop) services, thus its transit fare is high, due to the high operation cost. Public transits offer shared rides, thus reducing the cost of operations when there are a significant number of people riding together, say, on a bus. However, existing public transits operate along fixed routes with fixed time tables, where the transit capacity offered do not always match the time-varying trip demands. Consequently, many urban residents rely heavily on private cars and other transport modes (e.g., motor cycles, bikes) to get around a city, creating urban road congestion.

The aforementioned urban transport systems operate primarily in two modes: *fixed route mode* (with a large number of stops) in public transit services; and *point-to-point mode* in private passenger services. Differing from these two modes, *hub-and-spoke mode*[1] is a system of connections, where all traffic move along spokes connected through a small number of hubs. This mode has been extensively studied in the literature and is commonly used in industry, particularly in Airline route map planning [4], telecommunications [5], freight [2], and package delivery system. Hub-and-spoke mode has advantages over the other two transit modes in the following aspects: It requires less stops/transfers than existing public transits to save on trip time; it requires less routes than private transits, where the smaller number of routes may improve the efficiency of using transportation resources and increase the occupation rate.

In this work we propose *CityLines* (in analogy to "Airlines" for flight route services), a scalable dynamic hybrid hub-and-spoke transit system with shared shuttles. The CityLines service relies on a hybrid hub-and-spoke transit network, consisting of a set of interconnected hub stations in the urban area. Our main contributions are summarized as follows.

• We propose an optimal hybrid hub-and-spoke (OHHS) framework for CityLines system. A trip demand originated from a small region (referred to as a spoke region) is routed to the destination with a non-stop service (in the point-to-point mode) or via a hub station (in the hub-and-spoke mode). Given a city with n small regions (spokes), if a total budget allows L hubs and M point-to-point transit routes, CityLines selects the hub locations and assign urban trip

---

[1]*Fixed route mode* and *hub-and-spoke mode* both allow transfers during a trip, where fixed route mode relies on a large number of densely distributed stops/tranfers (e.g., one stop per kilometer) to serve passengers, and hub-and-spoke mode employs very few (usually less than three) hubs per trip to guarantee the quality of experience while aggregating trip demands.

demands to hubs or point-to-point routes to minimize the average travel time.

• We conduct experiments on real trajectory data of taxi, bus and subway collected during March 2014 in Shenzhen, China. The results demonstrate that CityLines provides a transformative urban transit service, with travel time as short as private transits and travel cost as low as public transits.

2 OVERVIEW

In this section, we will motivate and define hybrid hub-and-spoke planning problem, detail the datasets we use.

2.1 System Design Trade-offs and Motivations

The choice of urban transit services from a passenger depends on the QoE and cost of the trip, where the QoE hinges on many potential factors, including in-vehicle time, level of inconvenience, etc [6], and the trip cost depends on the service operation cost. Private transit services in general offer high QoE, with low in-vehicle time and high level of convenience, but at a high cost of trip fare. On the other hand, by reducing the operation cost with ride-sharing, public transit services have a lower trip fare, but longer in-vehicle time. Hence, due to the fundamental trade-off between passengers' QoE and operation cost, private and public transit services are operated to meet one of the two aspects, respectively. The next question is *how we can develop a transit service to serve urban trip demands with travel time as short as taking private transits and trip fare as low as taking public transits?* In this paper, by utilizing the historical trip data from urban transportation systems in Shenzhen, we make the first attempt to develop CityLines, a hybrid hub-and-spoke transit model, that allows an integration of both hub-and-spoke mode (to aggregate trip demands with small number of hubs, thus reduce the operation cost) and point-to-point mode (to reduce the overall trip time, thus to maintain a high passengers' QoE).

2.2 Problem Definition

The increasing prevalence of sensors, mobile devices, and Automated Fare Collection (AFC) devices has led to an explosive increase of the scale of spatio-temporal data, including passenger *trip demands* as defined as follows.

DEFINITION 1 (TRIP DEMAND). *A trip demand of a passenger indicates the intent of a passenger to travel from a source location src to a destination location dst from a given starting time t, which can be represented as a triple $\langle src, dst, t \rangle$.*

Passenger trip demands can be obtained from various data sources. For example, the transaction data from AFC devices in buses and subway systems record passenger trip demands at the level of bus stops and subway stations. Taxi GPS trajectory data with occupation information include the trip demands for taxi trips. For urban trip demands, we consider two types of transit modes below, i.e., point-to-point mode and hub-and-spoke mode.

DEFINITION 2 (POINT-TO-POINT MODE). *With point-to-point mode, a trip demand is served through a direct (usually the shortest or least-cost) path from the source src to the destination dst.*

The urban area consists of small regions, where a trip demand may originate from or destine to. Each of such small regions is referred to as a *spoke*. Some regions, referred to as *hubs*, are deployed with transfer stations, that allow trips to detour at. Given all spoke and hub regions, a hub-and-spoke transit mode can be interpreted as follows.

DEFINITION 3 (HUB-AND-SPOKE MODE). *With hub-and-spoke mode, a trip demand $\langle src, dst, t \rangle$ is detoured through a small number of ℓ hubs, $h_1, \cdots, h_\ell$ (with $\ell \leq 3$ in general). Thus, the path taken for the trip is $\{src, h_1, \cdots, h_\ell, dst\}$, and each segment of the path is in general a direct (least-cost) transit.*

Note that the more hubs a trip demand takes, the lower QoE a passenger would receive. In Airlines route planning, one hub detour is commonly used for trip demands. In this paper, to guarantee a high QoE, we allow $\ell = 1$ hub for a trip demand, where our framework also works for cases with $\ell > 1$.

Ideally, for those source-destination location pairs with a large number of trip demands, e.g., commute trips between a residential area and a commercial/working area, point-to-point mode is preferred. On the other hand, for those source destination pairs with less trip demands, hub-and-spoke mode is more promising to aggregate trip demands and reduce the operation cost by leveraging economics of scale. To balance such trade-offs, we propose to investigate the hybrid hub-and-spoke planning problem.

Problem Definition. Given a set of n spokes (regions) in an urban area, a set of K trip demands, and a budget of M point-to-point transit routes and L hub stations to deploy, we aim to find the optimal L regions to deploy hub stations and optimal assignment of trip demands to either point-to-point transit or a hub to detour from, so that the average travel time of all trip demands is minimized.

2.3 Data Description

To tackle the problem defined above, two real datasets are employed, including (1) trip demand data; (2) road map data.

Trip demands data are extracted from large GPS trajectory dataset (from taxis) and AFC billing dataset (from buses and subway trains) collected from Shenzhen, China during March 2014. For trip demands from buses and subway trains, we extract their starting and ending stations from the AFC billing data as source and destination locations. For taxi trips, we extract the source and destination locations by the occupation indicator in taxi GPS data.

Road map data. In our study, we use the Google GeoCoding [1] to retrieve the bounding box of Shenzhen. The bounding box is defined between $22.45°$ to $22.70°$ in latitude and $113.75°$ to $114.30°$ in longitude. The covered area is about $1,300km^2$. Within such a boundary, Shenzhen road map data were obtained from openstreetmap.org

3 METHODOLOGY

Figure 1 presents our optimal hybrid hub-and-spoke (OHHS) framework for CityLines system, consisting of three stages: (1) map gridding, (2) trip demand aggregation, and (3) optimal hybrid hub-and-spoke (OHHS) planning.

3.1 Stage 1: Map Gridding

The road map is divided into equal grids with a side-length of 0.01 degree in latitude and longitude. Then, a filtering process is conducted to eliminate those grids off the road network, so that the
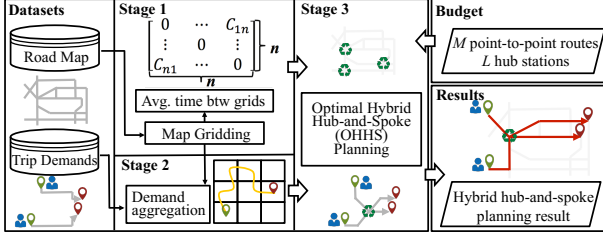
Figure 1: CityLines Framework

remaining n grids are strongly connected by the road map, namely, each grid can reach any other grid through the road map. We refer to those remaining grids as spokes in the urban area. Then, we estimate average travel time between each spoke pair. Thus, an n by n travel time matrix C is obtained, which contain the least travel time of each pair of spokes in the urban area.

3.2 Stage 2 :Trip Demand Aggregation

In this stage, all sources and destinations of trip demands are aggregated to the spokes extracted in stage 1. Hence, a trip demand $\langle src, dst, t \rangle$ is aggregated as $\langle s, s', t \rangle$, where s and s' are the spokes where source src and destination dst are located at. Then, a spoke level trip demand matrix V is obtained with each entry V_{ij} representing the number of trip demands originating from spoke i and terminating at spoke j.

3.3 Stage 3: Optimal Hybrid Hub-and-Spoke (OHHS) Planning

Consider a city with a budget of deploying point-to-point transit service for M spoke pairs, and L hubs for trip demands to detour. Given the spoke set G of n connected spokes, least travel time matrix $C = [C_{ij}]$, and volume matrix $V = [V_{ij}]$ as input, the hybrid hub-and-spoke planning problem aims to identify M spoke pairs to deploy point-to-point transit, L spokes from G to deploy hubs, and assign each of the rest source-destination spoke pairs to a hub for detour, so as to minimize the average travel time for all trip demands.

Without the point-to-point mode part, this problem is a well-studied combinatorial optimization problem, so called, p-HLP (p hub location problem), that aims to select a total of p hubs and assign each trip demand to one and only one hub, to minimize the average trip time. To include the point-to-point transit mode, we introduce a novel notion of *virtual hub*, denoted as h_0, which is not physically one entry from n hub candidates. Figure 2 illustrates how the virtual hub h_0 works. All trip demands assigned to h_0 are served by point-to-point transit mode. Instead, a trip demand assigned to a physical hub h_i ($1 \leq i \leq n$) will be detoured through h_i during the trip. By introducing the virtual hub h_0, the optimal hybrid hub-and-spoke (OHHS) problem can be formulated as follow.
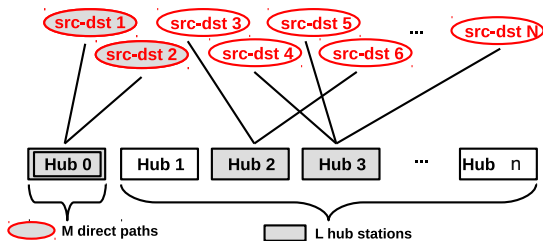
Let C_{ij}^k be the travel time for a trip demand from spoke g_i to g_j detoured at hub h_k. Recall that the least travel time from spoke g_i to g_j is C_{ij}. Thus, with a physical hub h_k, we have $C_{ij}^k = C_{ik} + C_{kj}$; and for the virtual hub h_0, we have $C_{ij}^0 = C_{ij}$, since a trip demand assigned to virtual hub h_0 is served with point-to-point transit mode. Let x_{ij}^k be a binary assignment variable indicating if trip demands with source-destination spoke pair (g_i, g_j) are assigned to hub h_k ($x_{ij}^k = 1$) or not ($x_{ij}^k = 0$). Moreover, we denote y_m (with $1 \leq m \leq n$) as a binary selection variable, indicating if a physical hub h_m is selected ($y_m = 1$) or not ($y_m = 0$). We want to resolve y_m, indicating the finally selected L hubs, and x_{ij}^k, the trip assignment to hubs, such that the average travel time of trip demands is minimized. This OHHS problem is presented below.

$$\text{min:} \quad \frac{1}{V} \sum_{g_i \in G} \sum_{g_j \in G} \sum_{0 \leq k \leq n} V_{ij} C_{ij}^k x_{ij}^k, \tag{1}$$

$$s.t.: \quad \sum_{0 \leq k \leq n} x_{ij}^k = 1, \qquad \forall g_i, g_j \in G, \tag{2}$$

$$\sum_{g_i \in G} \sum_{g_j \in G} x_{ij}^0 \leq M, \tag{3}$$

$$\sum_{g_i \in G} \sum_{g_j \in G} V_{ij} x_{ij}^k \leq F_k, \qquad 1 \leq k \leq n, \tag{4}$$

$$\sum_{1 \leq m \leq n} y_m \leq L, \tag{5}$$

$$y_m \geq x_{ij}^m, \qquad \forall g_i, g_j \in G, 1 \leq m \leq n, \tag{6}$$

$$x_{ij}^k \in \{0, 1\}, \forall g_i, g_j \in G, 0 \leq k \leq n. \tag{7}$$

$$y_m \in \{0, 1\}, 1 \leq m \leq n. \tag{8}$$

The objective function in eq.(1) indicates the average travel time of all trip demands, with $V = \sum_{g_i, g_j \in G} V_{ij}$ as the total number of trip demands to be planned. The constraint in eq.(2) states that each source-destination spoke pair should be served, i.e., by one and only one hub (including the virtual hub). The constraint in eq.(3) ensures that up to M source-destination pairs are served by point-to-point transit mode with direct paths. The constraint in eq.(4) specifies the capacity of each physical hub h_k, namely, the total number of trips going through a hub h_k cannot exceed the hub capacity F_k. The constraint in eq.(5) guarantees that the total number of physical hubs deployed is no more than L. Eq.(6) specifies a validity constraint, where a spoke pair (g_i, g_j) is assigned to a hub candidate h_m, if and only if h_m is selected to deploy a hub, namely, $y_m = 1$. The constraint eq.(7)–(8) indicate binary variables x_{ij}^k, y_m.

By introducing the virtual hub h_0 into the formulation, OHHS problem allows both hub-and-spoke and point-to-point modes. The nice property of OHHS formulation is that it still follows p-HLP (p hub location problem). In the literature, p-HLP has been extensively studied, with several efficient approximation approaches developed. For examples, Ernst and Krishnamoorthy introduced a 3-index formulation for p-HLP, which enables an LP relaxation based approximation solution [3]. Marin, Canovas and Landete introduced new formulations for p-HLP problem that generalized basic models with providing tighter LP bounds [7]. In this work, we adopt the solution proposed in [3] to solve our OHHS problem.
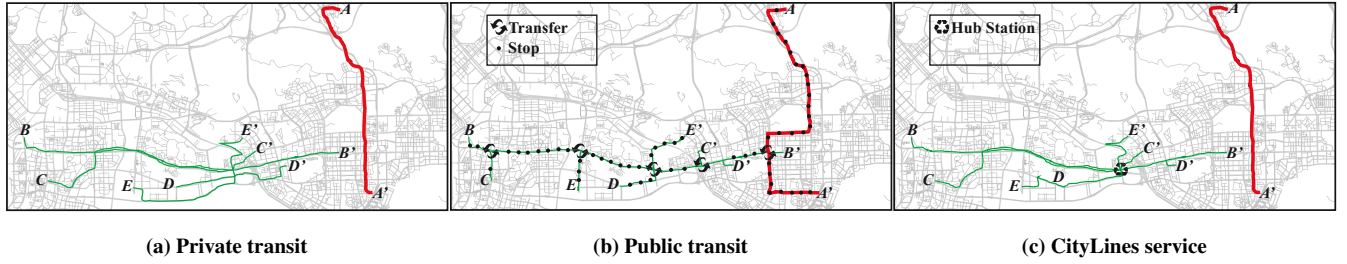

Figure 2: Illustration of the virtual hub

(a) Private transit | (b) Public transit | (c) CityLines service

Figure 3: Case Studies

4 EVALUATION

To evaluate the performances of our CityLines system, we conduct data-driven experiments using urban trip demand datasets collected from Shenzhen, China. Below, we elaborate on baseline methods, experiment settings and results.

4.1 Baseline Methods

We compare private and public transit models with our hybrid hub-and-spoke model.

(1) Private transit model: This model serves trip demands via direct least travel time paths with non-stop service.

(2) Public transit model: This model employs the existing public transit infrastructure (i.e., bus routes and subway lines), to serve all trip demands.

4.2 Experiment Settings

In the experiments, we evaluate operation cost using trip aggregation level, and evaluate the passenger QoE using average travel time.

Average travel time. Given a path planned for a trip demand $tr = \langle src, dst, t \rangle$ from the source to the destination, i.e., $\{g_1, \cdots, g_\ell\}$, the total travel time is given by $\sum_{2 \leq i \leq \ell} T_{i-1, i}$. The average travel time of all trip demands characterizes the quality of experience passengers receive from the planning strategy. The lower the time is, the higher QoE passengers experience.

Trip aggregation level (of trip demands). Given a planning method, each trip demand traverses a few *trip segments*. For example, in public transit model, the trips are divided into small trip segments between consecutive stop pairs. In CityLines service, each trip consists of spoke-to-hub and hub-to-spoke trip segments. In private transit model, each spoke pair maintains a unique trip segment as the direct path. Since trip demands may share the trip segments, each trip segment has a certain number of shared trip demands. The average number of shared trip demands per trip segment indicates the ride-sharing level, or trip aggregation level of the planning method. A higher trip aggregation level leads to lower operation cost.

4.3 Evaluation Results

Figure 3(a)–(c) show an example with real trip demands, which demonstrate the effectiveness of CityLines service by comparing it with private and public transit services. We extract a small set of trip demands during 6–11am in March 12, 2014, from Shenzhen, China. The trip demand set includes a total of 1,274 trip demands with 5 source and 5 destination spokes. One source-destination pair (from spoke A to A') is with the highest trip demand volume, i.e., 473 trip demands. Moreover, each source (from B, C, D) has

some trip demands (ranging within $58 - 118$) to each destination (in B', C', D'), and E has 77 trip demands to E'. Figure 3(a)–(c) show the trip planning solutions using three transit models, including private transit, public transit, and CityLines service (with one hub and one direct path as the budget). Our results show that private transit and CityLines lead to similar average travel time, as 23 and 26 minutes, respectively, and public transit has 47 minutes average travel time due to the large number of stops and transfers during the trips. On the other hand, public transit and CityLines enable similarly high aggregation levels, with 168 and 155 aggregated demands per trip segment, where private transit leads to only 112 aggregated demands[2], due to the distinct least travel time paths employed.

5 CONCLUSION

In this paper, we make the first attempt to develop CityLines system for urban transportation services, that employs a hybrid hub-and-spoke transit model. The model allows both point-to-point connection to improve the passenger quality of experience, and hub-and-spoke connection to reduce the system operation cost. The results demonstrate that CityLines framework achieves as short travel time as private transits and as high ride-sharing rate as public transits.

6 ACKNOWLEDGMENTS

REFERENCES

[1] Google GeoCoding. https://developers.google.com/maps/documentation/geocoding/.
[2] S. Çetiner, C. Sepil, and H. Süral. Hubbing and routing in postal delivery systems. *Annals of Operations Research*, 2010.
[3] A. T. Ernst and M. Krishnamoorthy. Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. *European Journal of Operational Research*, 1998.
[4] H. Karimi and M. Bashiri. Hub covering location problems with different coverage types. *Scientia Iranica*, 2011.
[5] H. Kim and M. E. O'Kelly. Reliable p-hub location problems in telecommunication networks. *Geographical Analysis*, 2009.
[6] C. Kumar, D. Basu, and B. Maitra. Modeling generalized cost of travel for rural bus users: a case study. *Journal of Public Transportation*, 2004.
[7] A. Marín, L. Cánovas, and M. Landete. New formulations for the uncapacitated multiple allocation hub location problem. *European Journal of Operational Research*, 2006.

---

[2]Note that the aggregation level of private transits is calculated without considering vehicle capacity. When using taxis, the aggregation level is up to 4, i.e., taxi capacity.