Improving Efficiency of Transit Systems in Mega-Scale Cities

Tian He
Department of Computer Science and Engineering, University of Minnesota
tianhe@cs.umn.edu

According to the Population Reference Bureau, 522 major cities in the world now have more than one million people as of 2013, among which 24 are megacities such as Tokyo, Delhi, New York and Shanghai with a population excess of ten million each. The raid expansion of Mega-scale cites leads to significant challenges in public transit systems. A recent study by Texas Transportation Institute [5] shows that in 2011, congestion caused urban Americans to travel 5.5 billion hours more and to purchase an extra 2.9 billion gallons of fuel for a congestion cost of $121 billion. It has been acknowledged that public transit systems played a significant role to prevent traffic situations becoming even worse than we are now. For example, researchers [3] show in 2007, public transit e.g., bus and subway reduced 640 million hours of passenger travel delay and 390 million gallons of gas in the U.S., achieving a saving of $13 billion congestion cost [2]. Taxicab system, a form of private transit system, also play a prominent role for residents' daily commute in mega-scale cities. Based on a recent survey in New York City [6], over 100 taxicab companies operate more than $13,000$ taxicabs with a stable delivery of $660,000$ passengers per day, and they transport more than 25% of all passengers, accounting for 45% of all transit fares paid.

Although urban transit systems operated and/or supervised the government have enhanced the efficiency of people's daily commute, there are a significant room for further improvements because of two major limitations. First, the existing public transit schedules are usually decided statically based on manually conducted surveys and census data [7], which are often out-of-date and incomplete. Second, the traditional theory and practice typically focus on small size isolated single transportation modes, and thus cannot efficiently collect, process, and utilize distributed, large-scale complex data sets about different modes of transportation, e.g., taxicabs, subway, buses, etc. These limitations result in poor passenger experience [4],*e.g.*, traffic congestion, unnecessary detours, long waking distances, high fares, prolonged travel times including waiting and riding times.

On the other hand, in recent years, infrastructures of intelligent transportation systems (ITS) in large cities have been upgraded with sensors and communication devices, such as widely adoption of intelligent fare automated fare collection (AFC) systems in bus, subway and taxi networks, massive amount steaming data about people's daily commute becomes available. For example, as the fifth largest metropolitan city with No.1 population density in China, Shenzhen (a city adjacent to HongKong) has developed a complex and advanced road network with 699 kilometers of highways. The city has deployed AFC device with GPS capability in 14,000 taxicabs, 10,000 buses, and 10,000 logistic vehicles. All 137 stations along 177 km subway network are equipped with AFC devices as well. Collaborating with the transport committee of Shenzhens municipality, our group and collaborators in China have obtained 98,472,628 of anonymized passenger pickup events and 3.9 billion GPS records in a 14,000 taxi network within six months of field study. We also have secure access over 6 billion anonymized records of smart card transaction over the AFC systems. These unprecedented amount of multi-mode real-time information about urban transit systems, unfortunately, is severely under-exploited and currently used only to visualize aggregated passenger traffic passing through subway gates and the locations of taxis around the cities.

We note the traditional theory and practice typically focus on small size isolated single transportation modes, and thus cannot efficiently collect, process, and utilize such comprehensive big transportation data in an effective and coordinated manner. As a result, applications, correlations and interactions among different modes of transportation are not sufficiently exploited. In this statement, we argue that there is an urgent need to investigate how to improve the efficiency of Urban Transit Systems by utilizing massive multi-mode real-time information collected by AFC systems. More specifically, we propose to deal with three major inefficiencies in existing transit systems with an integrated information infrastructure for next-generation urban transit systems.

- In taxicab industry, a long standing challenge is how to reduce taxicab's mileage spent without a fare, *i.e.*, cruising mile. The current solution for this challenge usually requires the participation of the passengers. We propose to build a cruising reduction system, $pCruise$, for taxicab drivers to maximize their profits by finding the optimal route to pick up a passenger, thus reducing the cruising mile. Base on the AFC records about occupancy and location of other near taxicabs, a taxicab characterizes its cruising process with a *cruising graph*. When a taxicab becomes vacant and tries to find a passenger, *cruising graph* will provide the shortest cruising route with at least one expected available passengers for this taxicab. With the shortest cruising routes, taxicabs will significantly reduce theirs cruising miles.
- Carpooling has long held the promise of reducing gas consumption by decreasing mileage to deliver co-riders. Although *ad hoc* carpools already exist in the real world through private arrangements, little research on the topic has been done. We propose to design, implement a new carpool service, called $coRide$, in a large-scale taxicab network intended to reduce total mileage for less gas consumption. In coRide, given demands obtained through AFC and call-in requests from passengers, we seek an optimal delivery graph to achieve the minimum total mileage. Although this optimization is NP-hard by nature, our preliminary investigation indicates that a 2-factor approximation solution with linear complexity exists. To accommodating the need for dynamic co-ride scheduling, it is necessary to investigate online streaming demands and also to consider different real world constraints, e.g., passenger travel periods, number of available taxicabs, and taxicab capacities. To encourage individual passengers to participate, it is essential to investigate incentive mechanisms to balance the benefits between drivers and passengers.
- The last mile problem, *i.e.*, many passengers' destinations lay beyond a walking distance to a public transit station, creates a long-standing barrier to better utilization of urban transit systems. To tackle this problem, we propose $Feeder$, a real-time on-demand service to deliver passengers from public transit stations to their final destinations with a data driven approach. We plan to utilize historical and real-time streaming data collected by AFC systems, *e.g.*, public transit and taxicab networks to accomplish the following design steps. (i) partitioning an urban area into appropriately sized regions, (ii) selecting effective stops within every region to reduce the last mile distance, (iii) calculating optimal routes connecting the stops with the minimum distance, and (iv) computing efficient schedules to dispatch feeder vehicles to minimize the average travel time.

Since many cross-cutting design issues are better exposed when they are examined under different scenarios, we propose to exploit aforementioned three inefficiency individually in the early stage of the research. In the later stage, we would like to harvest the knowledge obtained to design and implement an integrated architecture that allows multi-dimensional data from a wide range of transportation subsystems to be collected, processed and utilized efficiently, systemically and seamlessly. Specifically, we propose a three level architecture to acquire, process, and utilize information from multi-mode transportation to support ITS applications as in Figure 1. This three level design suggests the horizontal view to build high performance real world applications. The traditional stand-alone closed transportation systems, such as taxi networks or bus networks, do not have such capacities. The plug-in design feature allow us to efficiently add more transportation modes (such as bicycles) or applications without redesigning the whole architecture.

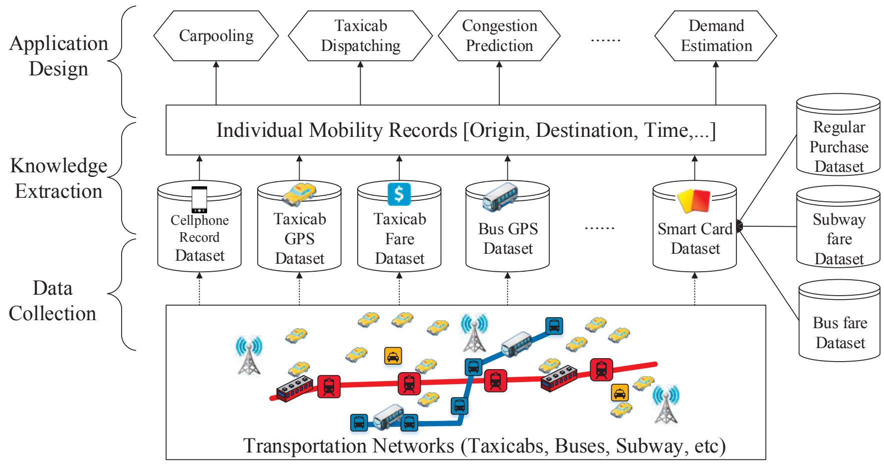

Fig. 1. Architecture Overview

– ***Data Collection*** requires reliable and seamless interconnection between AFCs in subway, taxi and bus networks and a central server cloud. In addition to existing collection schemes, we propose to design customized hardware to synchronize the data collection in heterogeneous AFCs. Our design features (i) low power sensing about physical environment and passenger behaviors, (ii) reliable collection model (Peer to Peer vs. Infrastructure Base), and (iii) feedback control based sensing synchronization.
– ***Knowledge Extraction*** unifies heterogeneous data from different modes of transportation to produce generic individual passenger mobility records, specifying origins, destinations and the other essential information. Our design aims to transparentize heterogenous features of different transportation datasets in terms of scale, timeliness, granularity and privacy, mining the high level knowledge to support various ITS applications. The dataset used to extract the passenger knowledge includes fare transaction data, vehicle location data, cellphone record data, etc.
– ***Application Design*** provides novel applications to enhance the passenger experience, *e.g.*, taxicab carpooling, taxicab dispatching, MiniBus service, demand estimation, *etc*. We propose to implement several applications in Shenzhen transportation system.

If successful, we will have a clear understanding on how to identify and address inefficiencies of existing transit systems in mega-scale cities. Effectively interconnecting different modes of transportation in metropolis-levels can reveal the useful correlation among different phenomena and cross the boundary of heterogeneous data, leading to novel ITS applications. Towards the very end, our work will fundamentally improve the quality of people's daily transportation in terms of cost, time and comfortability.

References Cited

1. Schaller Consulting. The new york city taxicab fact book. *http://www.schallerconsult.com/taxi/taxifb.pdf*.
2. T. Davis and M. Hale. Public transportations contribution to us greenhouse gas reduction. In *Technical report in Science Applications International Corporation*, 2007.
3. Brian Ferris, Kari Watkins, and Alan Borning. Onebusaway: Results from providing real-time arrival information for public transit. CHI '10, 2010.
4. Antti Nurminen and Juha Järvi. Real time reliability: Mixed reality public transportation. In *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, SA '13, 2013.
5. David Schrank, Bill Eisele, and Tim Lomax. 2012 urban mobility report. In *Texas AM Transportation Institute Report Series*, 2012.
6. NYC Taxi and Limousine Commission. Taxi of tomorrow survey. 2011.
7. Daisy Yoo, John Zimmerman, and Tad Hirsch. Probing bus stop for insights on transit co-design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, 2013.