TRANSIT HUB: A SMART DECISION SUPPORT SYSTEM FOR PUBLIC TRANSIT OPERATIONS

SHASHANK SHEKHAR¹, FANGZHOU SUN¹, ABHISHEK DUBEY¹, ANIRUDDHA GOKHALE¹, HIMANSHU NEEMA¹, MARTIN LEHOFER² AND DAN FREUDBERG³

¹ Institute for Software Integrated Systems, Vanderbilt University, Nashville, TN, USA
² Siemens Corporate Technology, Princeton, NJ, USA
³ Nashville Metropolitan Transport Authority, Nashville, TN, USA

36.1 INTRODUCTION

The allure of smart city technologies lies in its promise to enrich the lives of residents by empowering the stakeholders to make efficient and informed decisions and help alleviate complex issues. One such issue that is universally faced by large cities of the world is the problem of commuter traffic. Solutions are difficult, partly because so much of our physical infrastructure was designed for a different era—when no one could imagine just how many cars would hit the streets each day—and partly because options like mass transit and adopting car shares, biking, and walking are seen by many as “trading down” making their commute less convenient. The irony is that very often, a short walk or bus ride would be far faster, easier, and cheaper than driving. Unfortunately, there is a general lack of awareness among people about the effectiveness of these options.

For example, consider the city of Nashville, TN. The traffic congestion in the city has nearly doubled over the last decade and is expected to grow at an even faster rate in future—a recent study found it to be the top 25 congested cities in

© 2017 John Wiley & Sons, Inc. Published 2017 by John Wiley & Sons, Inc
Companion website: www.wiley.com/go/Geng/iot_data_analytics_handbook
the United States [1]. Therefore, there is an urgent need to act now. Given that
adding infrastructure is difficult and slow, we are focusing on increasing the use
of shared mobility options like public transit. Nashville ranks as 74 out of 100 in
the number of passengers per capita [2]. Improving transit services and increasing
their efficiency and usage are strategic priorities of the Nashville Metropolitan
Transport Authority (MTA). Key to attracting new users, in addition to providing
services that are competitive with personal automobiles and that can meet basic
service needs, is to make public transit easier to use and to improve the image of
the services. For those who seldom use public transportation, it can be difficult to
figure out how to interpret schedules and how to pay the fare, for example. Poten-
tial new customers have to be convinced that riding transit will be a pleasant
experience, and people like themselves use transit. However, without the ability
to model, it is almost impossible to efficiently design, configure, and deploy such
a system.

This lack of understanding is an impediment to improving and deploying of smart
city technologies at large scale. To that end, researchers from the Institute for
Software Integrated Systems at Vanderbilt University have teamed up with the
Nashville MTA and Siemens Corporate Technology to work on the Transit Hub [3]
project (Figure 36.1). This project aims to put accurate, real-time information about
potential travel options into citizens’ hands as soon as they choose their desired
destination from their current or a specified location.

Static schedule information for transit has been available in Nashville for a
number of years through tools such as Google Transit. However, vehicles do not
always exactly follow the schedule, making real-time vehicle locations and bus
stop ETAs valuable pieces of information for the customer. In order to provide
this type of real-time information, Nashville MTA has partnered with Trapeze
Group to design and install a fleet-wide vehicle tracking system called Automated
Vehicle Location (AVL). Transit Hub makes use of transit schedules in
combination with real-time vehicle location and service alert information from
the AVL system, as well as data from rider smartphones and dozens of other
sources. Among other analysis, anyone with our free app¹ can compare travel
times using all available options—biking, walking, public transit, driving,
and more.

As part of the larger smart city movement, we are implementing the Transit
Hub project as a large-scale distributed human cyber–physical system (CPS) [4],
wherein the human and the sensors in the physical environment provide the
information, which is analyzed by computational services to provide contextual
decision support to the commuters. The data collected by the system also helps us
create predictive analytical models that help the MTA better understand the transit
network bottlenecks and areas for future improvements. This article focuses on
the technical underpinnings of the Transit Hub application and the backend data
services.

¹The application is currently undergoing tests and is not yet available for wider dissemination.
36.2 CHALLENGES

One of the key challenges in enabling smart city applications such as Transit Hub is the ability to collect and disseminate sensor data from the selected element of the city infrastructure and then analyzing the data in almost real time. However, this is difficult because of scale and network heterogeneity. While we can mostly depend on the availability of cellular (e.g., 3G/4G and LTE) networks, we also have to contend with Wi-Fi LAN (e.g., DSRC/802.11p), WirelessHART, and Bluetooth technologies. Furthermore, all actionable information delivered to a nearby hub has a useful lifetime and must be disseminated prior to its temporal bound.

Our approach for solving the data dissemination problem relies on using modern middleware technology that can seamlessly allow configuration and management of distributed application across heterogeneous networks. We are currently investigating the application of a middleware developed by our lab—Android Mobile Middleware Objects (AMMO) [5], a novel middleware developed for tactical applications on mobile devices by Vanderbilt University for the DARPA Transformative Apps program. The goals of AMMO were to support multiple communication/data distribution paradigms over a diverse suite of networking technologies—from low-bandwidth tactical radios to cellular technology and Wi-Fi. Communications paradigms, such as broadcast publish–subscribe with content filtering, and client–server are supported. While developed primarily for tactical applications, the capabilities of AMMO are an excellent match for resource-constrained dynamic networks typical of a city-wide deployment. For example, smartphones with AMMO middleware were used for situational awareness during the presidential inauguration in January 2013 by first responders with the National Guard, Washington DC Metro Police, National Park Service, and the local Fire Department.

Another challenge that requires attention is the architectural framework for the decision support system. A practical solution is to use a multimodal simulation approach that facilitates the precise integration of heterogeneous, multimodel simulations. Simulation integration frameworks, such as High-Level Architecture (HLA), address the integration of distributed heterogeneous simulators using distributed discrete-event semantics. In this regard, we are building upon our prior work on Command and Control Wind Tunnel (C2WT) [6] a simulation integration framework that facilitates assessment of command and control systems performance in the presence of adversarial network disruptions.

36.3 INTEGRATED SENSORS

The Nashville MTA AVL system includes a number of integrated components and sensors to aid in the tracking, monitoring, and operating of transit buses. The primary sensor supporting vehicle tracking is a GPS antenna and receiver located onboard each vehicle. The GPS coordinates are sent to an onboard computer where the location information is then correlated with schedule and stop location data to determine where the bus is on the current route and trip. This location information is also used to calculate schedule adherence (minutes ahead of or behind schedule) for the bus.
In addition to the GPS receiver, there are additional integrated sensors to aid in the tracking of vehicles. For example, vehicle odometer readings are used to supplement GPS information in areas where signals are weak or unreliable. At Music City Center (MCC), the downtown transit hub for MTA, radio-frequency identification (RFID) transmitters have been installed at each bus bay. Every vehicle has an RFID tag that reads the signal from the MCC bay tags to determine when the bus arrives and departs the transit center. This equipment also allows transit supervisors to monitor the location of buses within the facility down to the individual bay level.

Other in-vehicle components are also integrated with the AVL system. The farebox receives route and location information from the AVL, enabling the correlation of fare payment activity by bus stop. Engine alarms are sent directly to the AVL by the Engine Controller Module to enable real-time vehicle health monitoring and historical data analysis. Infrared-based Automated Passenger Counter (APC) sensors installed at each bus door provide passenger boarding and alighting activity data to the AVL which then correlates the data with schedule information to determine passenger activity by stop, trip, route, and direction. Even the vehicle destination sign is integrated with the AVL to enable the sign to change automatically and without driver intervention when switching between routes.

36.4 TRANSIT HUB SYSTEM WITH MOBILE APPS AND SMART KIOSKS

The Transit Hub system is accessed by a mobile application that can be deployed on individual user’s smartphones. It features smart trip planning, service-alert integration, personal transit schedule management and notification, and real-time transit tracking and navigation. The Trip Planner utilizes origin and destination address and departure and arrival time to search for the future transit trips that meet the user’s requirements.

The Trip Planner offers a user-friendly interface (Figure 36.1) for users to enter the information for route searching. For people who are not familiar with the Nashville area, for example, first-time visitors, they can enter the start and end address in the search bar. For the local residents, they can just drag the map view to pinpoint the start and end locations.

The app’s real-time view (Figure 36.2) displays a map of the scheduled trip with lines indicating the walking/bus route and markers indicating the bus stops to transfer as well as original and destination locations. Users can tap the Go button at the very bottom to start real-time navigation. If real-time data is available for the trip, the exact bus that the user is supposed to get on/off for the next stop will be shown in the map, with time label indicating the remaining time. If there is no real-time data for the scheduled bus, time left to get on/off the next bus will be calculated and displayed based on the static schedule time.

The application not only aggregates the real-time transit data that is already available from local transit authorities, but it also provides a crowd-sourced alternative to official transit tracking feed. When a user is using the mobile application, they can choose to provide their anonymous location and the data about when they are on the bus to the backend server. The server uses anonymous data to protect user’s privacy.
FIGURE 36.2 Real-time view. Reproduced with permission from Abhishek Dubey.

As the data accumulates over time, techniques like transit simulation and machine learning can be utilized to process the huge database of collected data, predict transit delay with more accuracy, and make the transit service more reliable by adjusting transit schedules. Furthermore, it could even provide potential bus stops to add and remove to better meet the evolving demand patterns.

Notification is usually an important part of any transit application. Without proper notifications, users can still miss the buses if they forget to keep an eye on the departure time after searching the suggested routes and planning the ideal trips to destinations. In the Transit Hub application, users can enable push notifications easily from the app and customize how long they want to leave alerts to be pushed ahead of time. When a user schedules a trip into the future and adds it to the Calendar, the notification is set to be pushed to the user’s device when it is time to leave to catch the scheduled bus. Moreover, if the backend server predicts that the scheduled trip might be delayed for some reason, the user is notified accordingly.

36.4.1 Transit Hub Information Architecture

Figure 36.3 illustrates the technical design of the overall Transit Hub decision support system which enables the smartphone application. At the center lies the hub middleware responsible for coordinating all the Transit Hub activities. Its roles include
running data collection service, analyzing the collected feed, running simulations to provide a decision support framework in response to client requests. The data collection service is responsible for collating data from different sources and persisting into the Transit Hub’s distributed database for consumption by the decision support system and its clients. The service adheres to the timeliness and data quality guarantees needed by the decision support system. The service is also replicated across multiple servers in master–slave architecture such that one of the slaves take over the data collection role in case the master service fails. There are three types of data being collected:

1. Real-time feeds from Nashville MTA—Google has defined the General Transit Feed Specification (GTFS) [7] for transit authorities to release transit data feeds. The Nashville MTA publishes real-time feeds using the GTFS format. Transit Hub collects this data for supporting mobile applications and performing historical analysis for the city-level decision support system.

2. Traffic-related feeds—There are several sources of traffic-related information in Nashville city. The Tennessee Department of Transportation (TDOT)
publishes traffic-related feeds. HERE API [8] is another source for traffic congestion information. In the current iteration of the Transit Hub framework, we have used the HERE APIs as the data source. In the future, we will also include weather data and event details from social media channels.

3. Trip information from Transit Hub mobile application—The Transit Hub mobile application is another source of data for the decision support system. Based on user permissions, it can transmit the GPS coordinates of the rider’s bus trip, as well as the path traversed by foot. This is valuable information that can be utilized to plan future bus routes and their frequency to better utilize the available resources. Presently, we are only logging the searches performed by the users while planning their itinerary and their selection.

The multisource data is collected and persisted into different data stores in the Transit Hub backend for later retrieval. We employ MongoDB, a distributed NoSQL database to manage and store the data which is accumulating rapidly over time:

1. Real-time transit data from Nashville MTA: Our backend server repeatedly requests and stores the data from the trip updates feed, vehicle position feed, and service alerts feed every minute. The size of the data being stored in the database is about 3 GB per day.

2. Real-time traffic flow information from HERE API: We are recording the traffic flow data for road segments of all bus routes in Nashville for analysis and prediction purposes. Without optimization, the scale of the raw data is about 2.8 GB per day, which is extremely space consuming. To optimize performance and save storage space, we remove the static fields from original data such as road segment, physical layout information, and speed limits. Furthermore, since the traffic condition typically remains the same in a short term, we adopt a time series format that only stores the traffic condition which changes since last update. The optimized traffic data for storing is reduced to nearly 10% of its original size, about 0.27 GB per day.

3. Static bus schedule dataset: This dataset is updated only when Nashville MTA releases new bus schedules to the public.

4. Crowd-sourced data from Transit Hub app: Our backend server collects these data anonymously upon user permissions when users plan for bus routes or track their trips in the app. The size of the data depends on the quantity of mobile app users and how frequently they use the app.

36.4.2 Decision Support Framework for Transit Hub

The Transit Hub mobile application helps users in trip planning and real-time tracking. It relies on the Transit Hub middleware for supplying optimal trip plan options and better trip tracking than what is provided by the MTA feeds. To achieve this, Transit Hub performs analysis and provides decisions at two levels. At the global level, it integrates historically collected data with simulations to provide augmented
feeds to the mobile application and trip recommendations that helps MTA to optimize the load (future work). At the user level, the user’s travel history and current information, such as planned trip, location, time, and cost constraints are utilized to provide best recommendations for the user—this is part of the ongoing work.

36.4.2.1 Analyzing the Collected Data Feed  In the current version of Transit Hub, we provide augmented feeds to external entities and the Transit Hub mobile applications containing predicted time for different bus routes. The analytical engine of Transit Hub consists of a simulation-based predictive model, a data-driven statistical model, and a real-time prediction model.

The simulation-based model works with the real-time feed and current traffic delay information to simulate bus movement on various routes and predict delays. This model uses the Simulation of Urban MOBility (SUMO) [9] microscopic simulator for simulating city traffic. We use an OpenStreetMap (OSM) [10] of Nashville and convert it in the format that SUMO understands. The static routes for buses from MTA GTFS feeds are mapped to SUMO format for simulating. We maintain a pool of virtual machines to run simulations for different buses from their current locations. The real-time traffic conditions are obtained using HERE APIs. Going forward, we will also take into account the historical data from traffic congestion at different times of the day. The traffic congestion information contains a “jam factor” that provides information about congestion level at all road segments within the queried region. Based on the jam factor, we configure the simulator for lane speeds and periodically run multiple simulations to collect the result to augment the feeds with delay results produced in the simulation.

The statistical model applies long-term analytics methods on historical transit data to explore persistent delay patterns in route segments and bus stops. This model utilizes K-means clustering algorithm to group the historical delay data into different clusters according to the time in the day and performs normality test on each clusters. The analytics results, including the mean value with confidence interval of each cluster, are then provided to the analytics dashboard and Transit Hub mobile app. The model also helps to identify the outliers in real time. By investigating the bus trips with severe delay, we can understand better what factors cause bus delay and how to optimize the system. Our backend server runs the statistical methods for all the route segments at the end of each month when a new monthly historical dataset is ready.

The real-time prediction model utilizes the real-time transit feeds such as trip updates and vehicle locations to provide short-term delay prediction. Since the data rate of the original real-time vehicle location feed is not stable and there may be errors and delay in hardware and communication, we developed a Kalman filter to reduce the noise when estimating the arrival time at each bus stop. To predict the delay in a route segment, we utilize not only the predicted route’s data but also the data from multiple other routes that share the same route segments. Another Kalman filter is applied on the preceding trips’ delay data to predict the delay for the requested route. We will integrate more data sources into the system, such as weather conditions, traffic congestions, special events, and so on.
36.4.2.2 Dashboard and Recommendation Engine for City Planners

The Transit Hub decision support system will also cater to the needs of MTA engineers and city planners. The current implementation of the Transit Hub analytics dashboard is shown in Figure 36.4. Choosing the options from different routes, directions, weekdays, and time period in the day, the analysts from MTA and the city can utilize the information provided by the dashboard to check the historical delay patterns and outliers in each route segment, identify which parts are the bottleneck in the routes, and come up with solutions to modify the bus schedules and optimize the performance of the transit system.

It will enable what-if analysis based on simulations and historical traffic and demand data and help in answering questions like how do the riders get affected if the number of buses are reduced or increased, buses get rerouted or a traffic accident occurs? It will also assist in simulating cascaded delays due to congestion. Designing incentives for passengers to take buses from particular stops and routes based on simulated results is another goal.

36.4.2.3 Infrastructure Requirements for Transit Hub

The Transit Hub has been designed to form the backbone of Nashville MTA’s scheduling and planning services. This imposes strict requirements on the system that needs to be fault tolerant and provide timeliness guarantees. The users of Transit Hub application also expect availability and timeliness between service level agreements (SLAs). The scale of the system brings its own challenges.

The real-time feeds accumulated by the data collection service are at the scale of several gigabytes per day. This number will keep increasing as MTA expands its services and we introduce new data sources. This requires us to design infrastructure that can efficiently persist and query data at terabyte scale and handle petabyte scale data in the longer run. We also need data replication to withstand failures. To fulfill these requirements, we use a distributed NoSQL database, MongoDB residing on a cluster of servers. However, going forward, we need a database which can support even larger datasets.

Another resource-intensive component of Transit Hub is the Analytical Engine that periodically or on demand runs simulations to predict delays on different routes. However, maintaining a huge pool of virtual machines to perform simulations is expensive. We need efficient resource management algorithms so that the service is viable. To that end, we are developing algorithms utilizing Linux container-based virtualization technology [11].

36.4.3 Kiosk Systems for Human–CPS Interaction

Today’s platforms have evolved from being simply technical innovations to also focus on realizing powerful, socioeconomic platforms affecting the daily lives of millions of users. These socioeconomic platforms enable a new class of applications and services to emerge at an unprecedented speed, quality, and cost. These trends can help cities and urban areas to increase their pace of innovation, while the effort expended in planning and implementation can be reduced from time spans ranging
from years or even a decade to just months. As an example, consider the state of art for user interfaces or wearable interfaces: the smartphone is the common denominator.

Despite the proliferation of smartphones, both socioeconomically backward and senior citizens are unlikely to afford today’s sophisticated smartphones or be able to use them effectively. Consider public transportation as a basic civic amenity provided by a city. People having difficulties using smartphones are often the most reliant on public transportation, for example, elderly or disabled people. Often they need to rely on information from station personnel, security personnel, police officers, or fellow citizens. Despite several new mobile apps that help plan for transportation, people with lower socioeconomic status often lack the resources to access those services. Access to public transportation, including the information to plan trips has even been named a civil rights issue. These socioeconomic issues demonstrate that communities and transportation providers cannot rely solely on smartphones for providing transportation-related information to citizens. This is where the notion of a smart kiosk device comes into the picture.

Kiosk Systems enable new ways of interaction between humans and CPS. Touch-based computers and terminals have been placed in public settings for more than 20 years, but the advancements of hardware and software as well as novel architectural approaches such as CPS enable a completely new class of system. Kiosks are envisioned to be strategically placed in cities and urban communities to provide riders access to transportation-related information. Barrier-free access, both in terms of placement of the Kiosk and design of Kiosk hardware and software, is a key consideration. All users benefit from the additional effort spent on designing an accessible system. In this section, we will introduce Kiosk Systems and their benefits for cities as well as their citizens, visitors, and local business. We will also show how Kiosk Systems can complement apps such as the Transit Hub.

Kiosk Systems offer direct and barrier-free access to city-life-related information and services. They integrate real-time, contextual, event-aware, and localized content from various domains. They can themselves serve as an information source using modular extendable sensors and usage feedback. Figure 36.5 highlights one such Kiosk System, and Figure 36.6 summarizes the key benefits of Kiosk Systems and illustrates how they help in making smarter decisions.

36.4.3.1 Benefits for Cities Due to the advent of social media and the decline of traditional media, many cities and public agencies have adopted new means of communication with their citizens, visitors, and customers. However, cities and transportation agencies do not control when and where the information gets displayed. So, often this information drowns in a stream of tweets, emails, or posts. Also, often the information may not reach the intended audience in a timely manner. Even regular users of public transportation are not interested in service updates of routes they are not using or are not affecting their commute. Casual users will not subscribe to those updates at all, because the information might be relevant to them only a few times per year. All these challenges can be resolved by having the kiosks display the right information at the right time.
FIGURE 36.5 Example of a Kiosk System—Siemens smart city hub. Reproduced with permission from Abhishek Dubey.

FIGURE 36.6 Decision support by a Kiosk System. Reproduced with permission from Abhishek Dubey.

Additional benefits accrue from the ability of the kiosks to provide instant feedback to city officials because the kiosks can track and report on the likes and dislikes of citizens who utilize the services of the kiosks thereby enabling the officials to refine their offering. A prominent capability of the kiosk is the fact that it can be customized for the needs of the city. This capability stems from the understanding that cities have different needs, they face different challenges, they have different geographical features, and they may have different historical past. Having a kiosk be tailored to the needs of the city is a significant benefit for the city.
36.4.3.2 Benefits for Citizens  Kiosk Systems enable quick access to services such as public transport, bike sharing, parking, games, and navigation integrated into one interactive platform. The reason why this is more attractive is because this approach does not incur the same impediments as those incurred in traditional communication mechanisms, such as mass mailings or local newspapers. These latter approaches cannot convey real-time information and updates. The Kiosk System enables information on and booking of such services. The hardware platform itself facilitates easy information consumption by using large multitouch screens. Free (wireless) Internet access and charging facilities for smartphones or tablets are additional benefits for citizens. Charging facilities can also be extended to electric bicycles and cars. Many more capabilities can be included in a kiosk. For example, it can be equipped with medical equipment ranging from simple first-aid kits all the way to lifesaving defibrillators. They can also be used to track localized crime, such as an assault, and help dispatch security personnel to the scene in a timely manner.

36.4.3.3 Benefits for Entrepreneurs and Local Businesses  Kiosk System enable winning new customers by means of advertising that are not considered as such. If a traveler is looking for a small, independent café that is not located at a nearby location, a Kiosk can provide him or her directions to the café. Incentives such as coupons can also be targeted effectively, for example, limiting them to a certain period of time when the business is slow or showing them to customers who have not visited a business yet. Thus, kiosks have the potential to boost local economy. In turn, the customer service in these businesses is bound to improve further improving the reputation of the city.

36.5 CONCLUSIONS

As the urban sprawl increases across the world, the cities of the world are facing unprecedented challenges stemming from traffic congestion, housing costs, environmental pollution, water and sanitation issues among many others. Often many of these challenges cannot be effectively addressed due to the bureaucratic structuring of the city government and the lack of interoperability across various departments. Beyond these limitations, however, many of these challenges remain unresolved also because technology has not been harnessed to its fullest potential despite significant advances in wireless and mobile connectivity, proliferation of smart end devices such as smartphones and other IoT technologies, and powerful computational platforms such as cloud data centers.

This chapter described our efforts in harnessing these technological advances in the context of smart cities. Specifically, we have focused our efforts on the Transit Hub project being designed and deployed for the city of Nashville, Tennessee, USA. The Transit Hub is a multilayer architecture comprising the Nashville MTA's buses which act as the sensors of our IoT architecture and providing their location information, a trip planning mobile app running on smartphones that enables
REFERENCES

calculators to plan their trips, and a cloud-based decision support system that analyzes
real-time traffic and bus location data to serve the trip planning requests made by
smartphone users. Complementing the smartphone-based app is also the Smart Kiosk
system, which can be deployed at strategic locations within the city to guide visitors
and also city residents, particularly those from socioeconomically backward strata
who cannot afford smartphones or the elderly who cannot operate the smartphones,
in guiding them to their destinations.

Preliminary ideas and working artifacts from our Transit Hub project were demon-
strated at the 2015 Global City Teams Challenge program [12] hosted by the US
National Institute for Standards and Technology (NIST). Our team is now partnering
with the Nashville city government addressing multiple additional challenges beyond
just the bus trip planning. We aim to make further progress in the future and demon-
strate our ideas at venues such as GCTC 2016, as well as evaluate the efficacy of our
technology and its benefits to Nashville. Based on these outcomes, we plan to reach
out to other cities providing them insights gained from our work and make our tech-
nology available to them.

ACKNOWLEDGMENTS

This project uses derivatives from work funded in part by the US National Science
Foundation and Siemens Corporate Technology.

REFERENCES

bizjournals.com/nashville/blog/2014/06/report-nashvilles-traffic-congestion-33rd-worst-
MTA/docs/StrategicTransitMasterPlan/06Ch4PeerReview.pdf (Accessed September 28,
2015).
Karsai. 2012. Rapid synthesis of high-level architecture-based heterogeneous simulation:
September 28, 2015).
