Transit-Hub: Building a Multi-time scale analytics system

Abhishek Dubey, Jules White, Fangzhou Sun, Chinmaya Samal, Hiba Baroud

Vanderbilt University

In collaboration with

Martin Lehofer

Siemens, Corporate Technology, Princeton, NJ

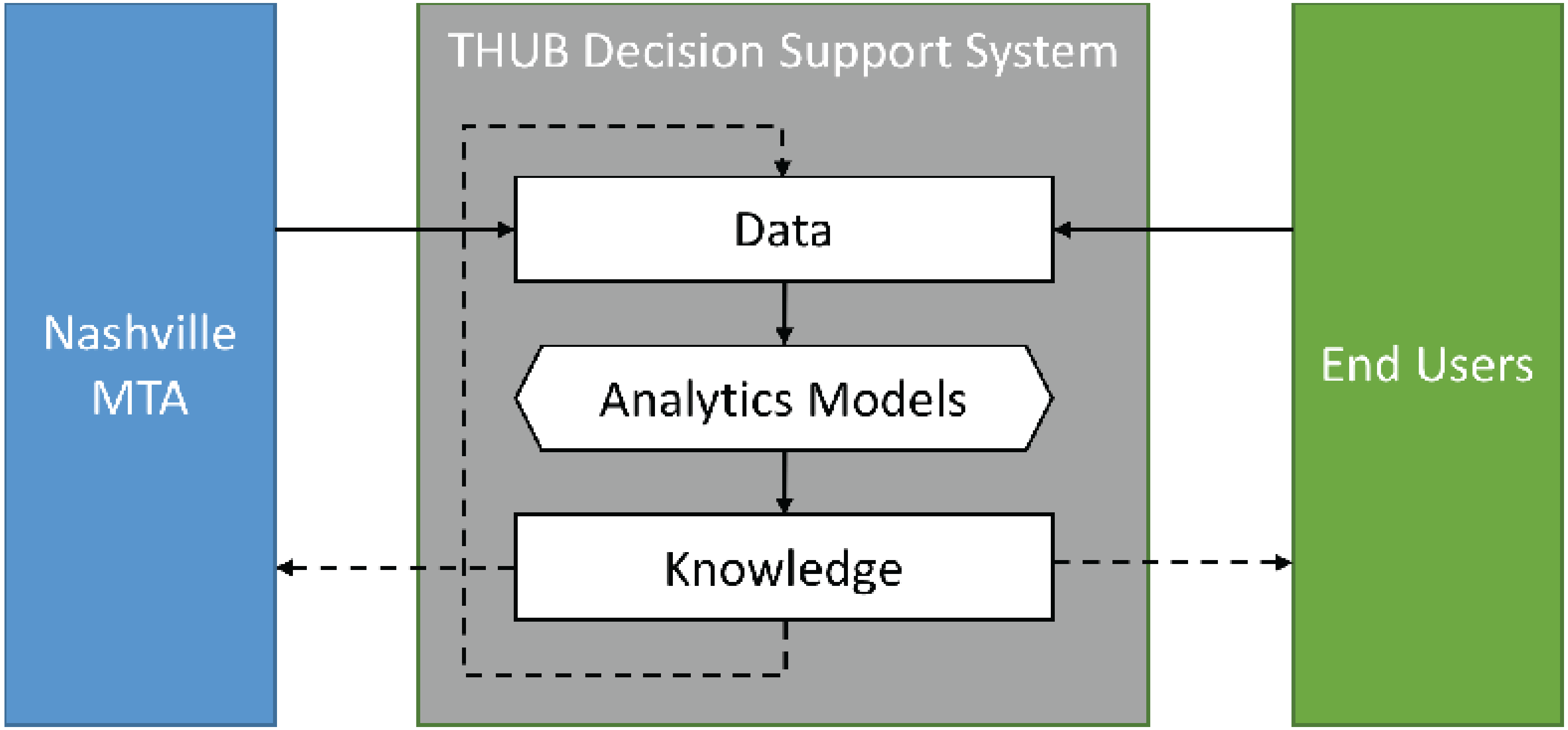
Supported by

CNS-1528799

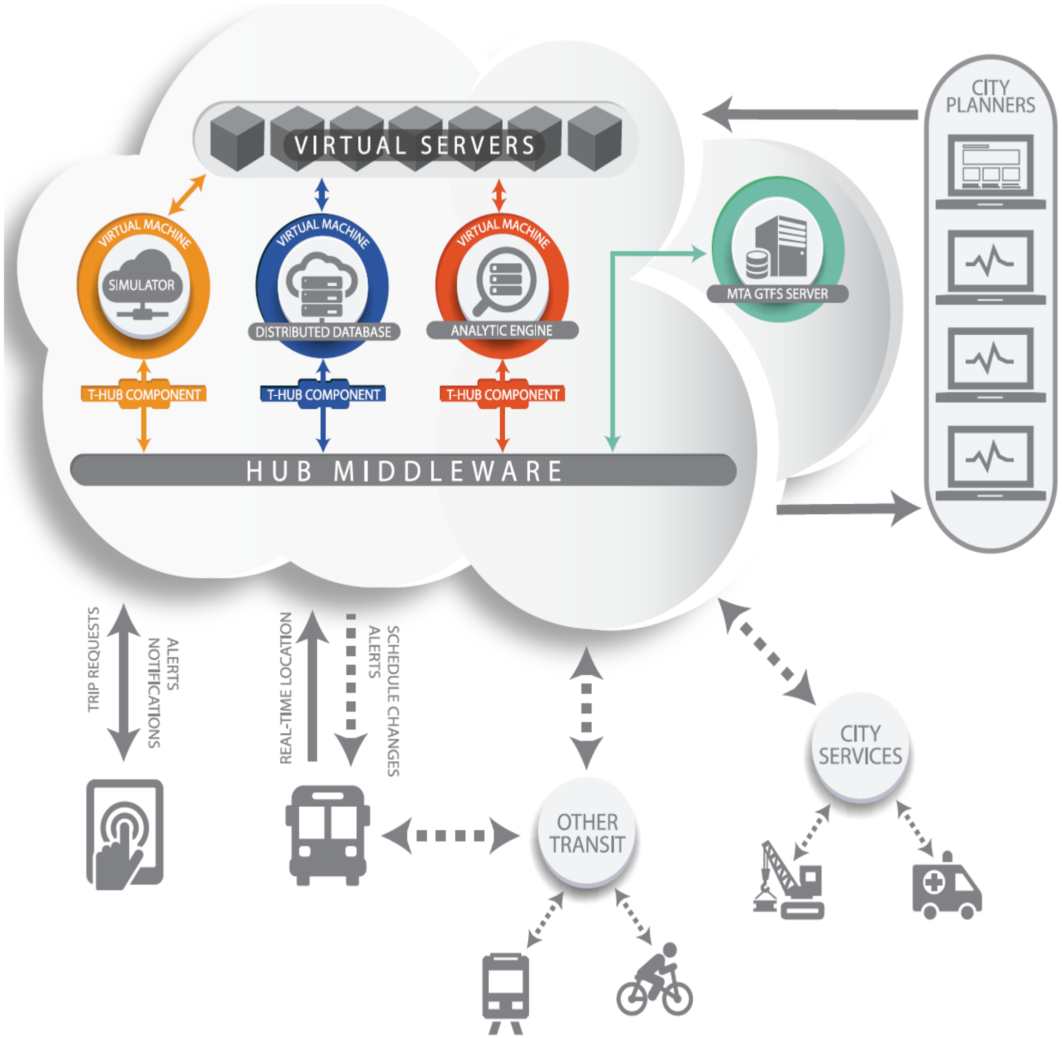
System of Interest

www.tennessean.com

Goal – Decision Support System



Approach – Transit-Hub



- **Architecture**
  - “Dynamic merge” of travel modes (walking to stop, public transport, connections) with real-time traffic simulations
  - Advanced analytics engines for cities as well as users
- **Real-time data ingestion**
  - Real-time feed of vehicle locations
  - Real-time traffic feed using Nokia HERE APIs
- **Advanced decision support**
  - Advanced trip planning
  - Notifications and alerts
  - Rescheduling
  - City services planning
- **Incentive-based ridership promotion**
  - Integrated health monitoring
  - Cost and gas savings
  - Carbon credit calculations
  - Integration with city incentives

Intellectual Challenges

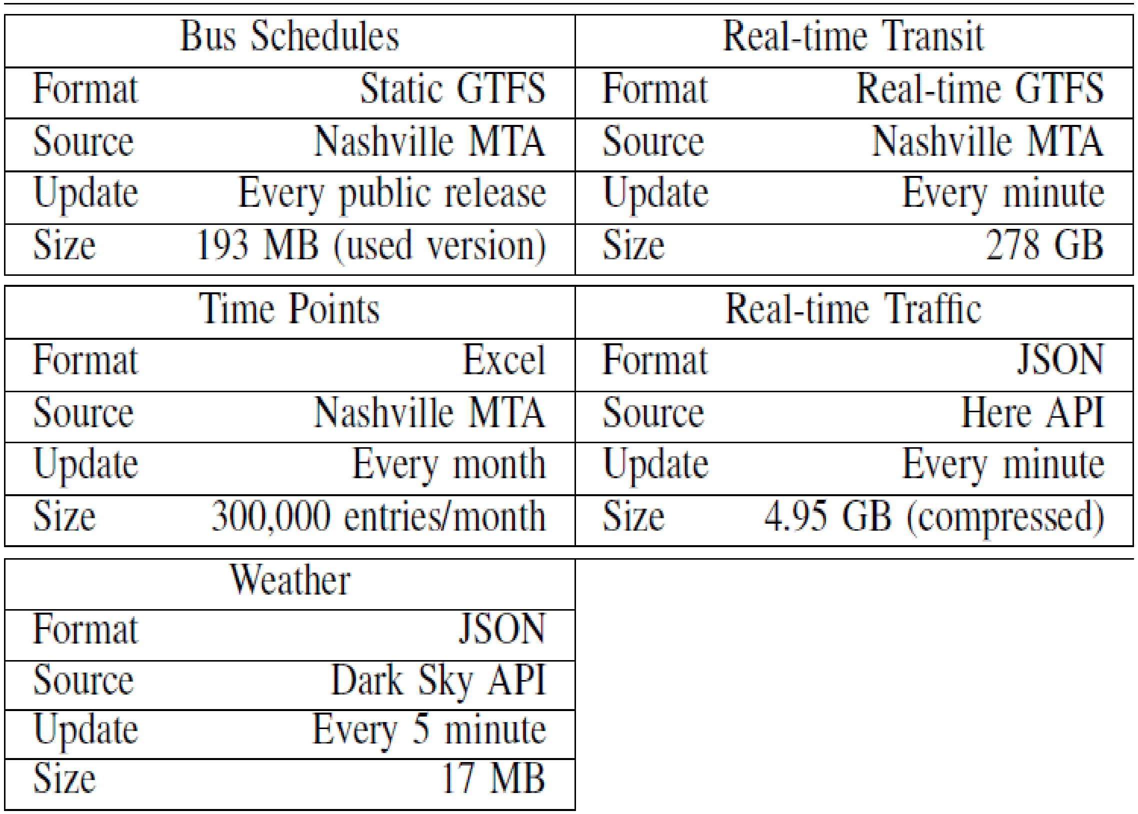
- Building online data mining and analytics for providing both real-time suggestions as well as improving the transit service in long-term.
- Increasing the user interaction mechanism beyond smart phone applications and providing a path for interoperating with third party services.
- Ensuring that we can distribute the computational components beyond traditional cloud while ensuring resilience and reliability.

Intellectual Challenges

- Building online data mining and analytics for providing both real-time suggestions as well as improving the transit service in long-term.
- Increasing the user interaction mechanism beyond smart phone applications and providing a path for interoperating with third party services.
- Ensuring that we can distribute the computational components beyond traditional cloud while ensuring resilience and reliability.

Delay Radar – Online Data Mining and Analytics

- We are developing techniques to combine multiple data sources for providing short term and long term delay prediction.



Bus Schedules		Real-time Transit	
Format	Static GTFS	Format	Real-time GTFS
Source	Nashville MTA	Source	Nashville MTA
Update	Every public release	Update	Every minute
Size	193 MB (used version)	Size	278 GB

Time Points		Real-time Traffic	
Format	Excel	Format	JSON
Source	Nashville MTA	Source	Here API
Update	Every month	Update	Every minute
Size	300,000 entries/month	Size	4.95 GB (compressed)

Weather	
Format	JSON
Source	Dark Sky API
Update	Every 5 minute
Size	17 MB

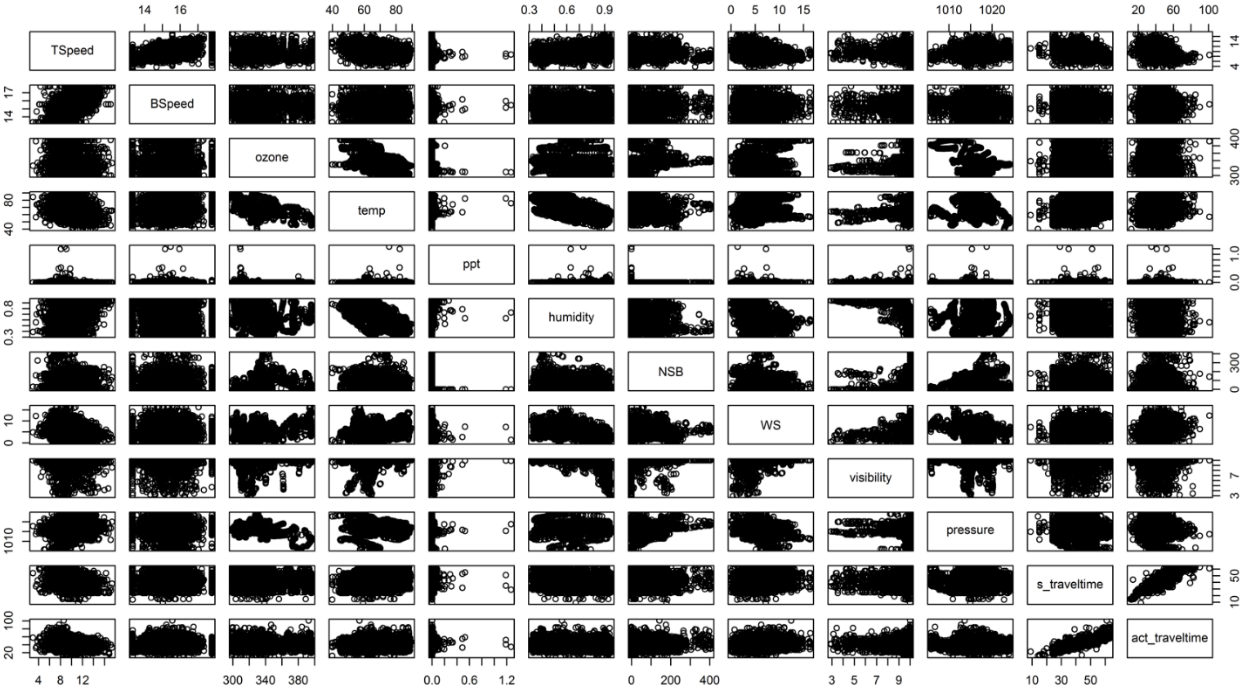
Table I

REALTIME AND STATIC DATASETS COLLECTED IN THE SYSTEM.

DelayRadar: A Multivariate Predictive Model for Transit Systems, IEEE Big Data, December 2016.

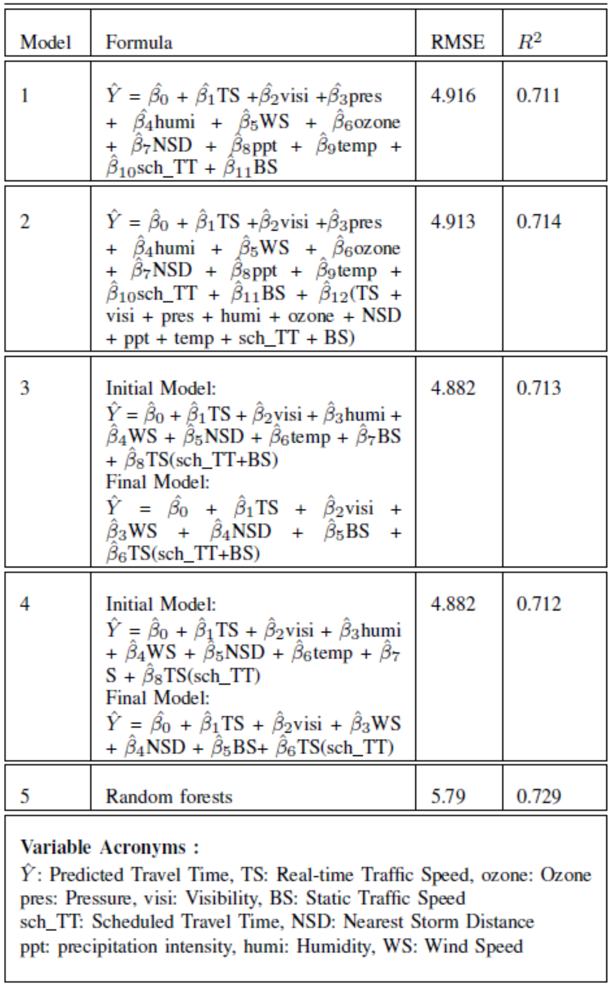
Delay Radar - Multivariate Predictive Models

- Predicting the effect of multivariate on transit system delay.
  - Weather condition
  - Traffic flow
- Long-term delay predictive model
  - Multivariate linear regression
  - Random forests
- Understanding the seasonal variations



Delay Radar - Multivariate Predictive Models

- Predicting the effect of multivariate on transit system delay.
    - Weather condition
    - Traffic flow
- Long-term delay predictive model
    - Multivariate linear regression
    - Random forests
- Understanding the seasonal variations

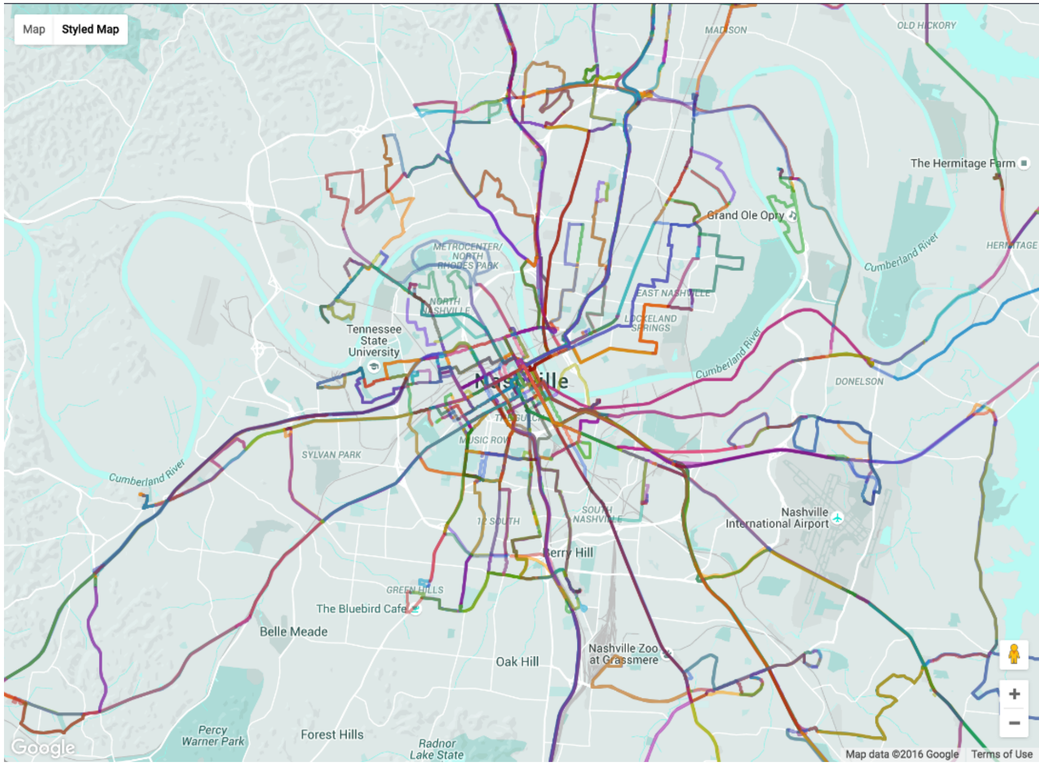


Model	Formula	RMSE	R^2
1	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\text{TS} + \hat{\beta}_2\text{visi} + \hat{\beta}_3\text{pres} + \hat{\beta}_4\text{humi} + \hat{\beta}_5\text{WS} + \hat{\beta}_6\text{ozone} + \hat{\beta}_7\text{NSD} + \hat{\beta}_8\text{ppt} + \hat{\beta}_9\text{temp} + \hat{\beta}_{10}\text{sch\_TT} + \hat{\beta}_{11}\text{BS}$	4.916	0.711
2	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\text{TS} + \hat{\beta}_2\text{visi} + \hat{\beta}_3\text{pres} + \hat{\beta}_4\text{humi} + \hat{\beta}_5\text{WS} + \hat{\beta}_6\text{ozone} + \hat{\beta}_7\text{NSD} + \hat{\beta}_8\text{ppt} + \hat{\beta}_9\text{temp} + \hat{\beta}_{10}\text{sch\_TT} + \hat{\beta}_{11}\text{BS} + \hat{\beta}_{12}(\text{TS} + \text{visi} + \text{pres} + \text{humi} + \text{ozone} + \text{NSD} + \text{ppt} + \text{temp} + \text{sch\_TT} + \text{BS})$	4.913	0.714
3	Initial Model: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\text{TS} + \hat{\beta}_2\text{visi} + \hat{\beta}_3\text{humi} + \hat{\beta}_4\text{WS} + \hat{\beta}_5\text{NSD} + \hat{\beta}_6\text{temp} + \hat{\beta}_7\text{BS} + \hat{\beta}_8\text{TS(sch\_TT+BS)}$ Final Model: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\text{TS} + \hat{\beta}_2\text{visi} + \hat{\beta}_3\text{WS} + \hat{\beta}_4\text{NSD} + \hat{\beta}_5\text{BS} + \hat{\beta}_6\text{TS(sch\_TT+BS)}$	4.882	0.713
4	Initial Model: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\text{TS} + \hat{\beta}_2\text{visi} + \hat{\beta}_3\text{humi} + \hat{\beta}_4\text{WS} + \hat{\beta}_5\text{NSD} + \hat{\beta}_6\text{temp} + \hat{\beta}_7\text{S} + \hat{\beta}_8\text{TS(sch\_TT)}$ Final Model: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\text{TS} + \hat{\beta}_2\text{visi} + \hat{\beta}_3\text{WS} + \hat{\beta}_4\text{NSD} + \hat{\beta}_5\text{BS} + \hat{\beta}_6\text{TS(sch\_TT)}$	4.882	0.712
5	Random forests	5.79	0.729

Variable Acronyms :
$\hat{Y}$: Predicted Travel Time, TS: Real-time Traffic Speed, ozone: Ozone
pres: Pressure, visi: Visibility, BS: Static Traffic Speed
sch_TT: Scheduled Travel Time, NSD: Nearest Storm Distance
ppt: precipitation intensity, humi: Humidity, WS: Wind Speed

Delay Radar- Core Algorithm

- Compute the route segments from the given transit schedule and map
- Get the recent delay data from routes that share the segment
- Calculate delay patterns:
  - Compute time intervals that have the highest similarity of delay patterns.
  - The mean of the time interval that is closest to the current prediction as the current delay is used as the current delay for the route segment.
  - We filter the delay via a smoothing filter afterwards.
- The delay over the whole route is computed by analyzing the delay across all the route segments.

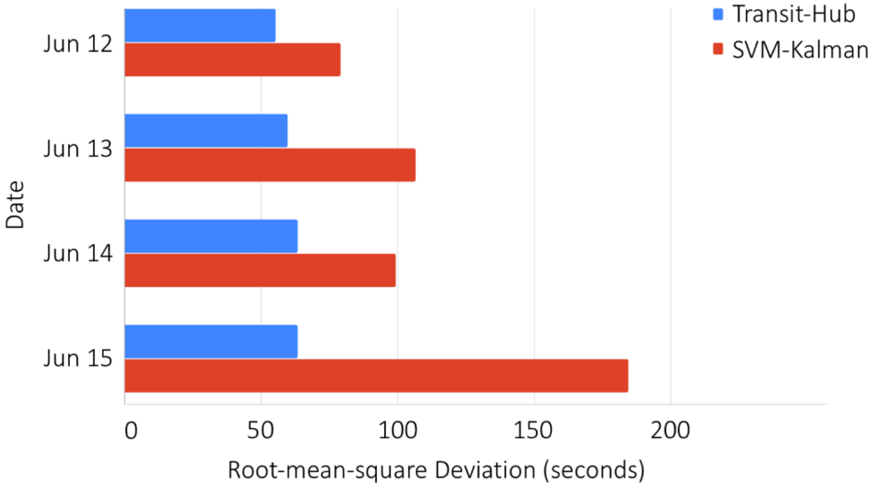


- 57 bus routes
- 5139 shared segments

The result is the travel delay prediction for the segment and hence the whole route

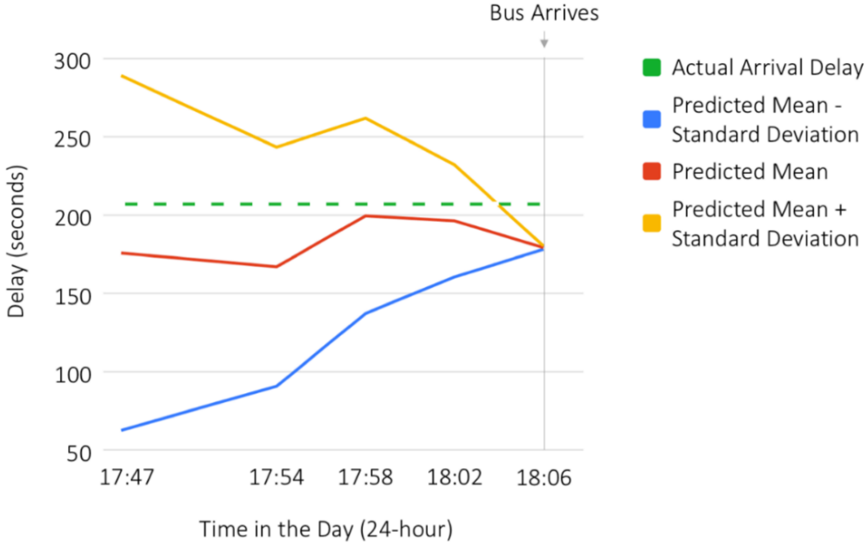
Delay Radar Algorithm

Experiment 1:



RMSD of travel time delay prediction when comparing the Transit-Hub model with the SVM Kalman model proposed in 2015

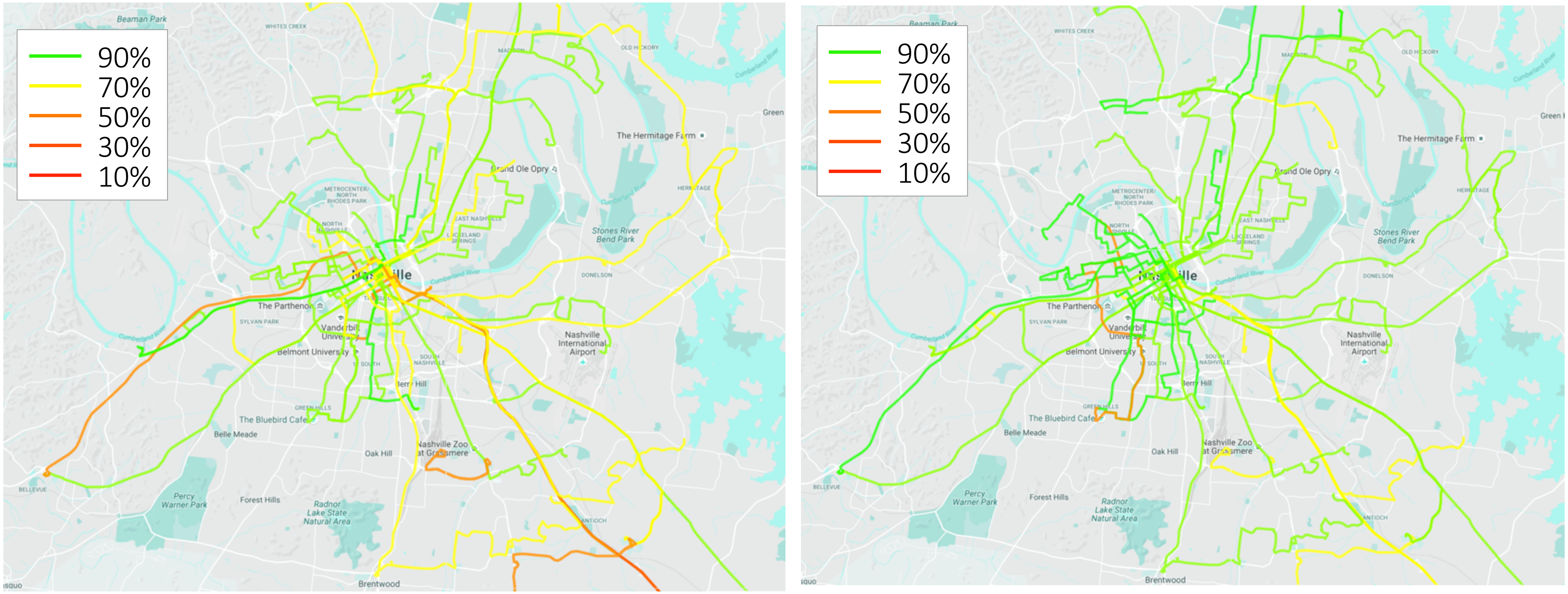
Experiment 2:



Arrival delay prediction for Route 3 and a specific stop.
The graph shows how the confidence interval (90%) bound tightens over time.

Bai, Cong, et al. "Dynamic bus travel time prediction models on road with multiple bus routes." *Computational intelligence and neuroscience* 2015 (2015): 63.

Delay Radar Scheduling Feedback



Before

After (Simulation)

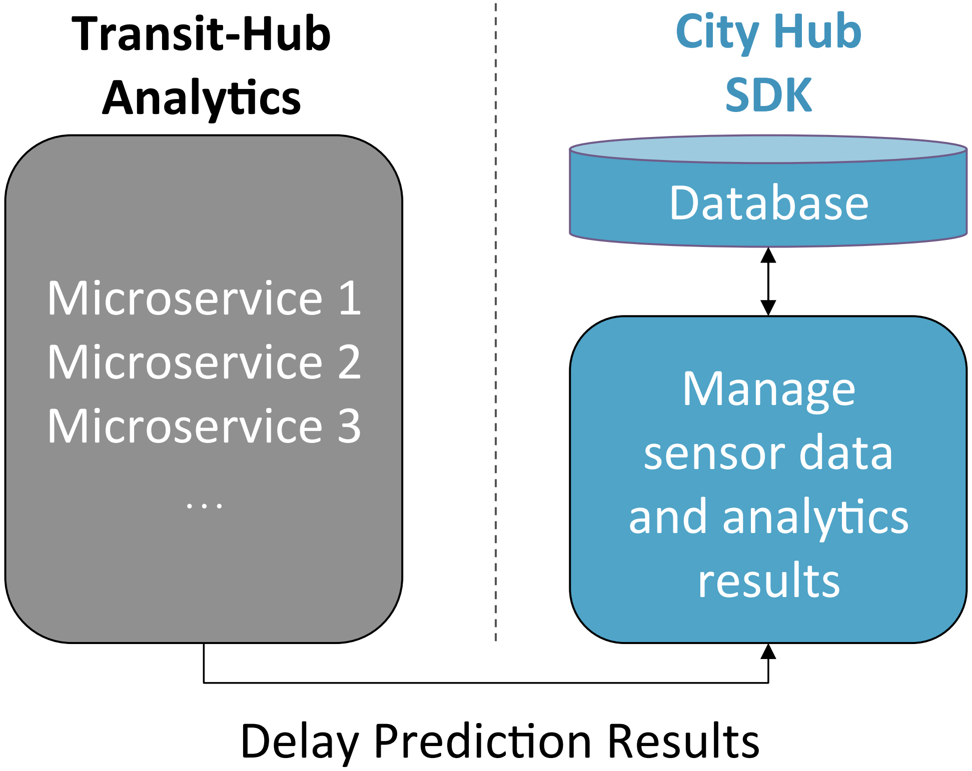
Showing percentages of historical trips in June 2016 where bus arrival delay at time points is between [-6 min, +1 min]

The changed schedule is currently under consideration at MTA

Intellectual Challenges

- Building online data mining and analytics for providing both real-time suggestions as well as improving the transit service in long-term.
- Increasing the user interaction mechanism beyond smart phone applications and providing a path for interoperating with third party services.
- Ensuring that we can distribute the computational components beyond traditional cloud while ensuring resilience and reliability.

Siemens City Hub - Novel user interaction mechanism



Transit-Hub Analytics

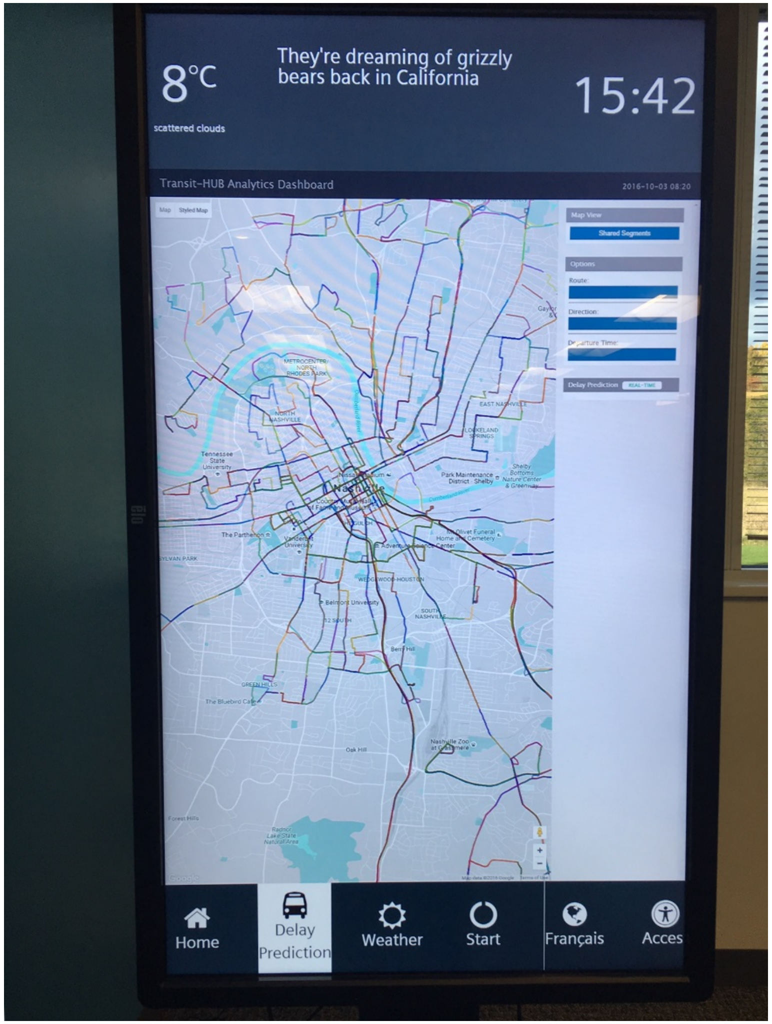
Microservice 1
Microservice 2
Microservice 3
…

City Hub SDK

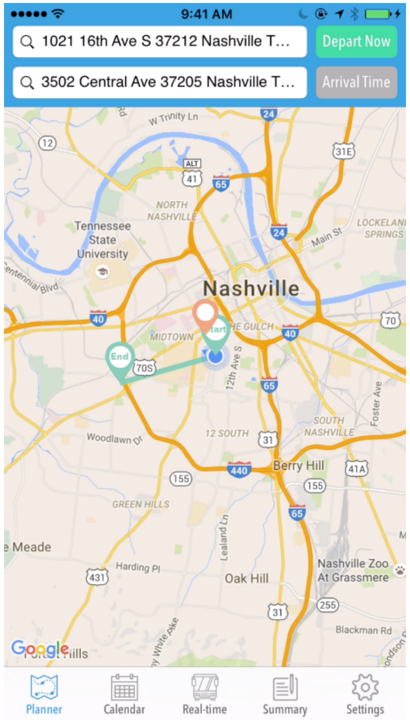
Database

Manage sensor data and analytics results

Delay Prediction Results



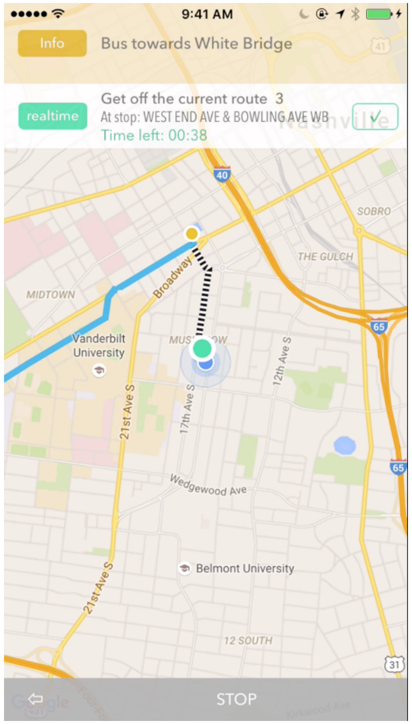
Transit Hub – Integrated Application



1021 16th Ave S 37212 Nashville T…

3502 Central Ave 37205 Nashville T…

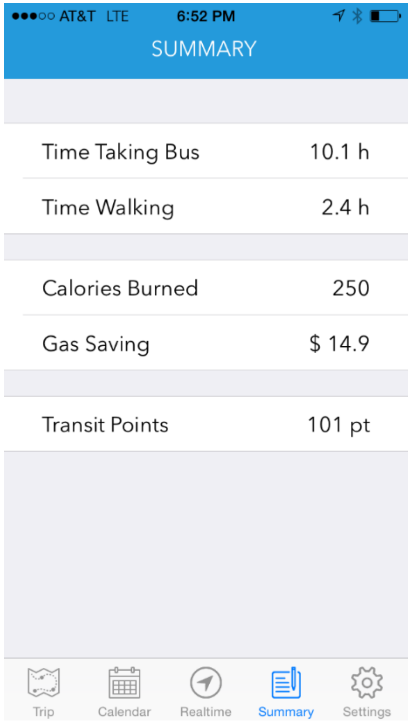
Trip Planning



Bus towards White Bridge

Get off the current route 3
At stop: WEST END AVE & BOWLING AVE WB
Time left: 00:38

Real-Time View

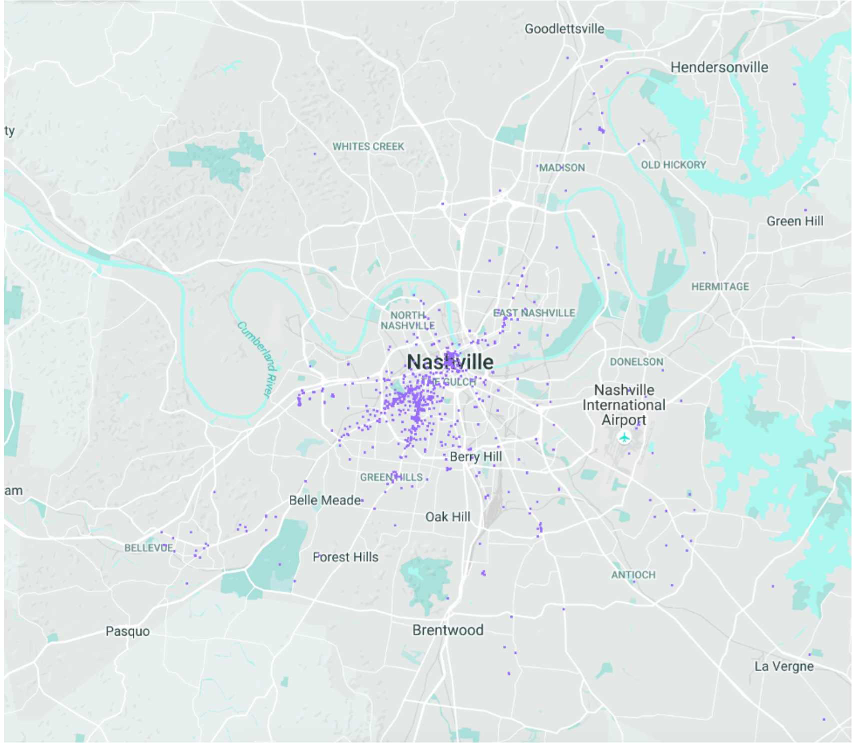


Time Taking Bus	10.1 h
Time Walking	2.4 h
Calories Burned	250
Gas Saving	$ 14.9
Transit Points	101 pt

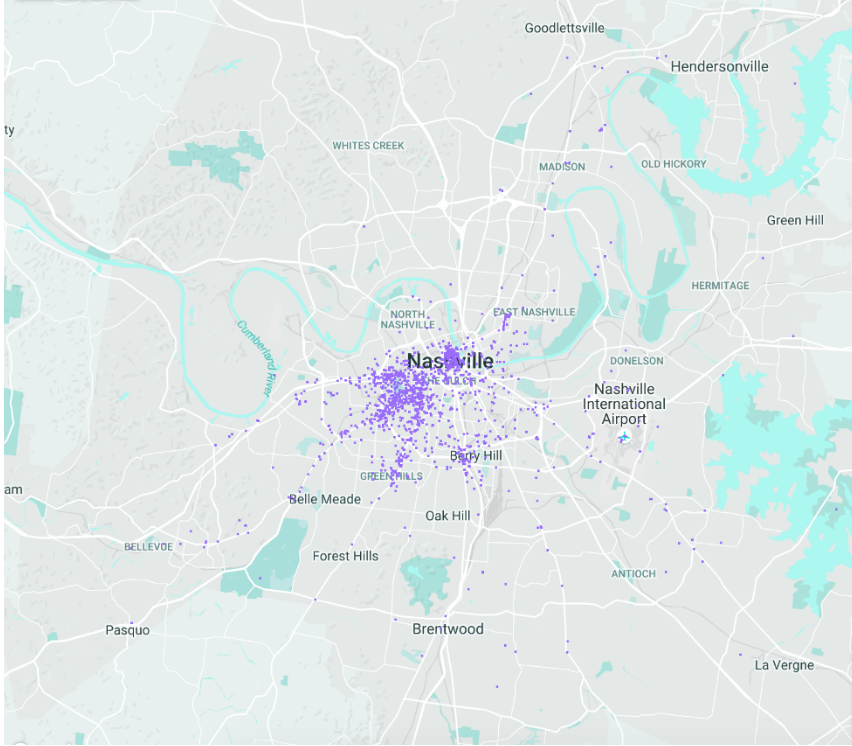
Incentives

User Activity

Destination addresses are more widely distributed across the city, while the departure addresses are concentrated along main roads.



Departure

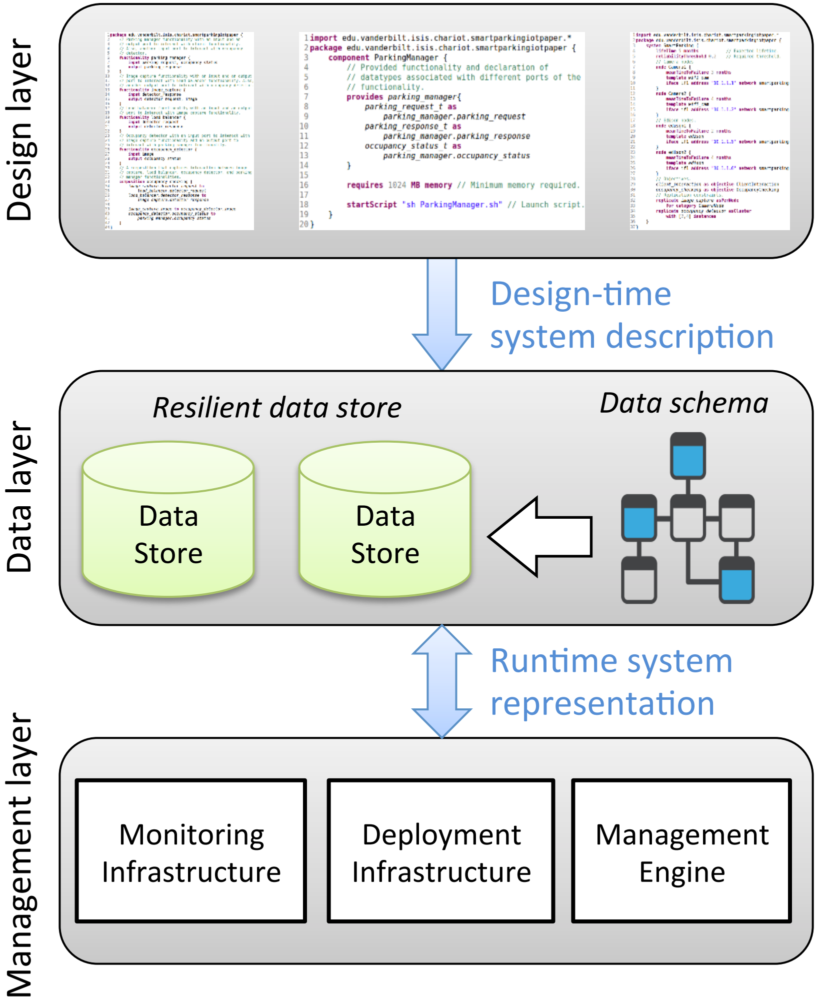


Destination

Intellectual Challenges

- Building online data mining and analytics for providing both real-time suggestions as well as improving the transit service in long-term.
- Increasing the user interaction mechanism beyond smart phone applications and providing a path for interoperating with third party services.
- Ensuring that we can distribute the computational components beyond traditional cloud while ensuring resilience and reliability.

CHARIOT Overview



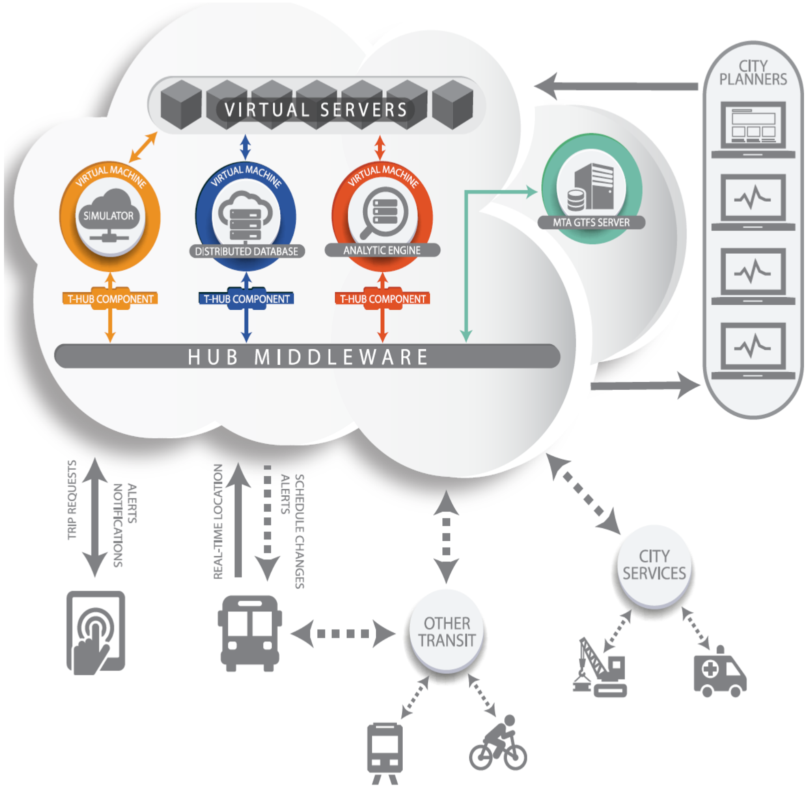
```
import edu.vanderbilt.isis.chariot.smartparkingiotpaper.*
package edu.vanderbilt.isis.chariot.smartparkingiotpaper {
    component ParkingManager {
        // Provided functionality and declaration of
        // datatypes associated with different ports of the
        // functionality.
        provides parking_manager{
            parking_request_t as
                parking_manager.parking_request
            parking_response_t as
                parking_manager.parking_response
            occupancy_status_t as
                parking_manager.occupancy_status
        }

        requires 1024 MB memory // Minimum memory required.

        startScript 'sh ParkingManager.sh" // Launch script.
    }
}
```

- CHARIOT (**C**yber-p**H**ysical **A**pplication a**R**ch**I**tecture with **O**bjective-based reconfigura**T**ion) is a holistic solution for managing extensible CPS

- CHARIOT comprises of entities that can be classified into three layers:
  - Design layer: DSML (Domain-Specific Modeling Language) to model systems, and interpreters to retrieve information from system models
  - Data layer: Codifies the format in which system information (comprises design-time system description and runtime system representation) must be stored
  - Management layer: Facilitates autonomous resilience by implementing a self-reconfiguration mechanism (sense-plan-act)
    - Failure avoidance
    - Failure management
    - Operations management

Future Extension

- US-Ignite: Social Computing Platform for Multi-modal transportation optimization
  - In collaboration with Baosen Zhang and Lillian Ratliff from University of Washington.
  - The solutions will be deployed in Nashville as well as Seattle



Intellectual challenge – peer to peer computing for distributed optimization

Can we lose the cloud?

Summary

Social Challenge:

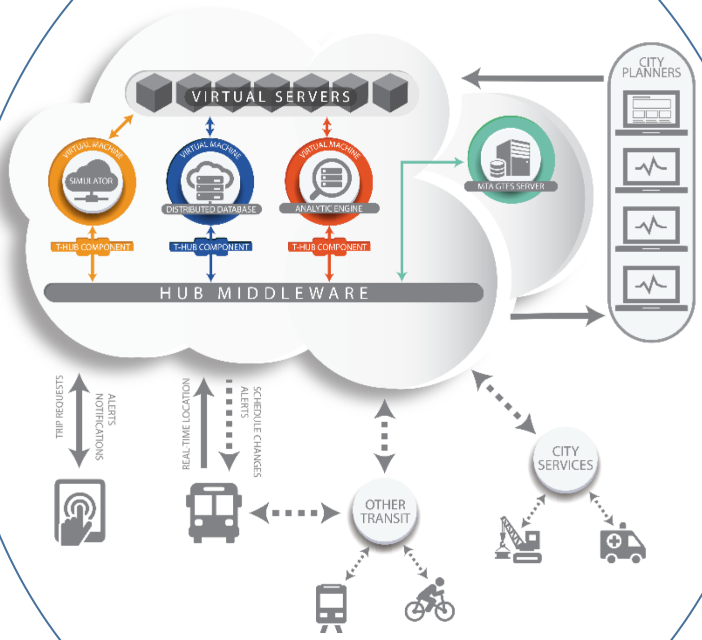
- Public Transit System is currently significantly underutilized in Nashville.
- Making transit services easier to use is important

Technical Challenge:

- Data-Driven Decision Support System
- Integrating heterogeneous sources of data
- Making the service extensible

Solution:

- City-Hub Integrated Application
  - Real-time planning
  - Service Alert Integration
- A decision support system
  - Data-driven analytics
  - Integrated Simulation based approach for "what-if" analysis
- Middleware for managing the distributed system



Scientific Impact:

- An extensible middleware framework that can be used to integrate other city services e.g. parking management
- Framework to combine historical data-driven analytics with real-time analytics.
- Understanding how decision support systems and incentive campaigns affect human engagement with the system
- Overall 6 publications from the project to date

Broader Impact:

- Deployment in Nashville.
- Inclusion of undergraduates in the research.
- Use of transit-hub as one of the project in a multidisciplinary university
- Initial research results have encouraged us to set up a center for focusing on smart city research.
- The next phase is focusing on multi-modal transportation.

Abhishek Dubey, Jules White,
Sandeep Neema
Award # 1528799