Herbert Wertheim College of Engineering
UNIVERSITY of FLORIDA

POWERING THE NEW ENGINEER TO TRANSFORM THE FUTURE

Department of Civil and Coastal Engineering

Real-time Urban Traffic Monitoring By Using Transit Buses As Probes

Shangkun Jiang, Yuran Sun, Wai Wong*, Yiming Xu, Xilei Zhao

01/10/2024

Research Motivation

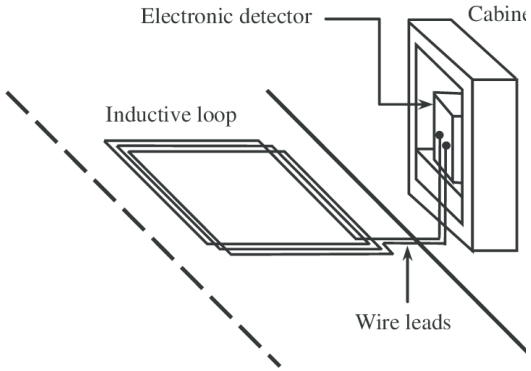
- Real-time urban traffic monitoring is a critical component in ITS and smart city development.
- However, challenges persist in immediate access to the real-time data.
  - Cost
  - Privacy
  - Coverage
  - Accuracy
  - Accessibility
  - …

Source: https://blog.ptvgroup.com/en/city-and-mobility/big-data-mobility/

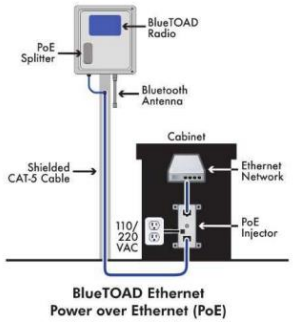
Source: https://datafromsky.com/news/datafromsky-real-time-traffic-monitoring/

Research Motivation

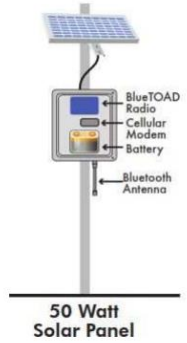
On-road fixed detectors:

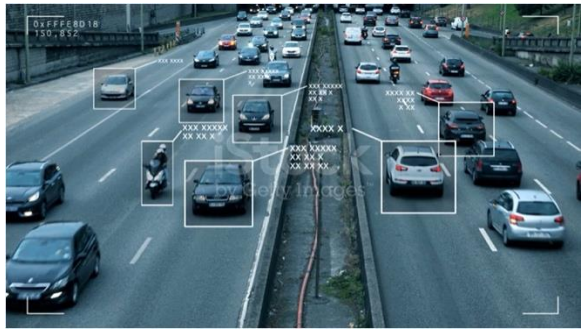


Loop detectors



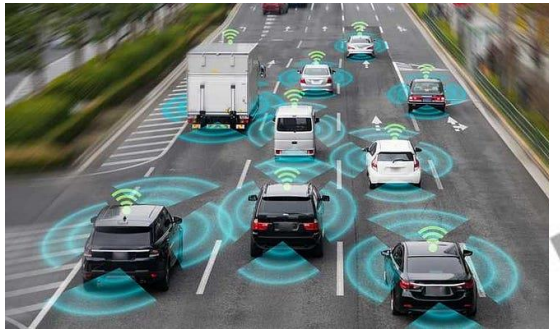
Bluetooth



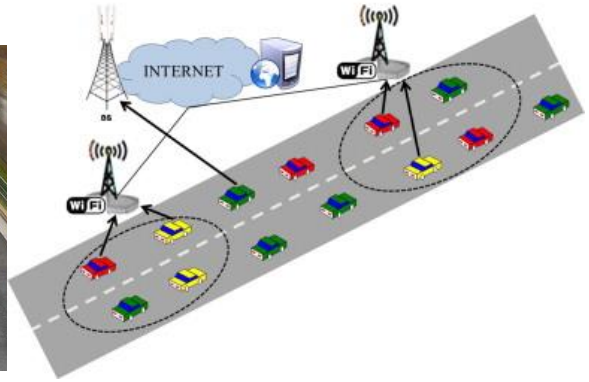


Video camera

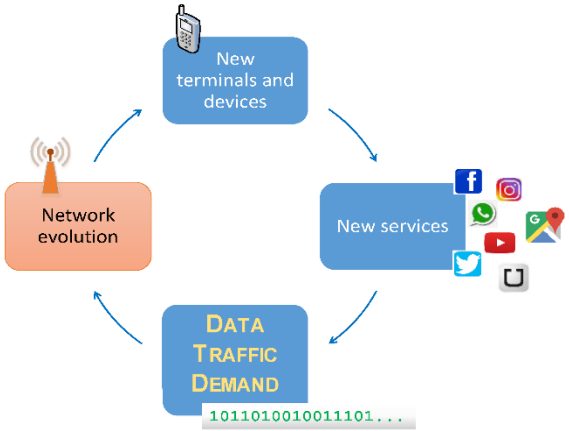
Mobile data sources:



Connected vehicles



Floating cars



Location-based service data

Research Motivation

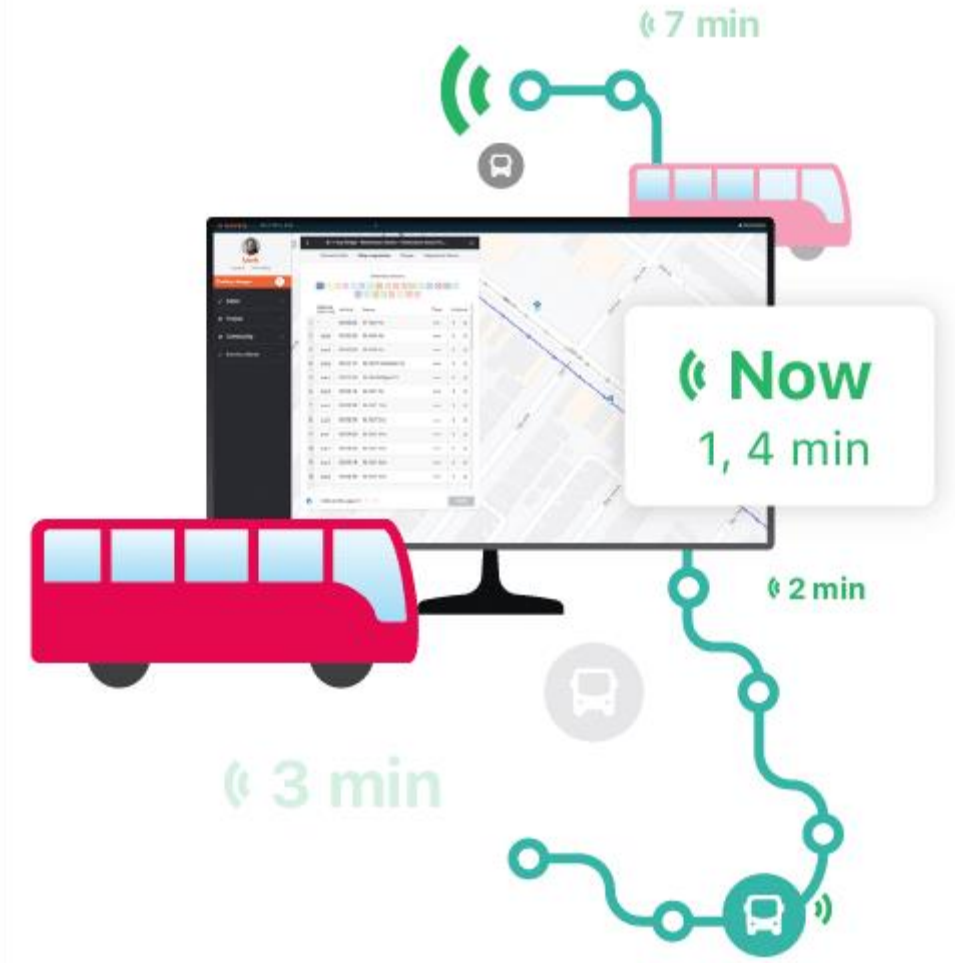
TABLE 1: Comparison of the pros and cons for datasets used for urban traffic monitoring.

Performance	On-Road Fixed Detectors			Mobile Data Sources				
	Loop detectors	Traffic camera	Traffic sensors	floating car data	CV data	LBS data	Google traffic data	GTFS data
Cost Effectiveness	*	*	*	*	*	*	*	***
Accuracy	*	**	*	**	***	*	*	***
Temporal Contiguity	**	*	**	**	**	*	***	**
Pre-Processing Simplicity	*	*	*	*	*	*	***	***
Spatial Coverage	*	*	*	***	***	**	***	**
Accessibility	*	*	*	*	*	*	**	***
Privacy Protection	***	*	**	*	*	*	**	***
Penetration Rate	/	/	/	*	*	**	***	**

Notes: The number of "*" represents the performance of the grading variables;
Cost Effectiveness: the overall cost-effectiveness of collecting data;
Accuracy: the degree of different data sources precisely reflects the real-world situation;
Temporal Contiguity: time continuity of the different data sources;
Pre-processing Simplicity: the straightforwardness of processing the raw data;
Spatial Coverage: the geographical spatial coverage of different data sources;
Accessibility: the easiness of data acquisition;
Privacy Protection: the protection of travelers' private information across different data sources;
Penetration Rate: the market share of data sources in real-time urban traffic monitoring.

GTFS data description

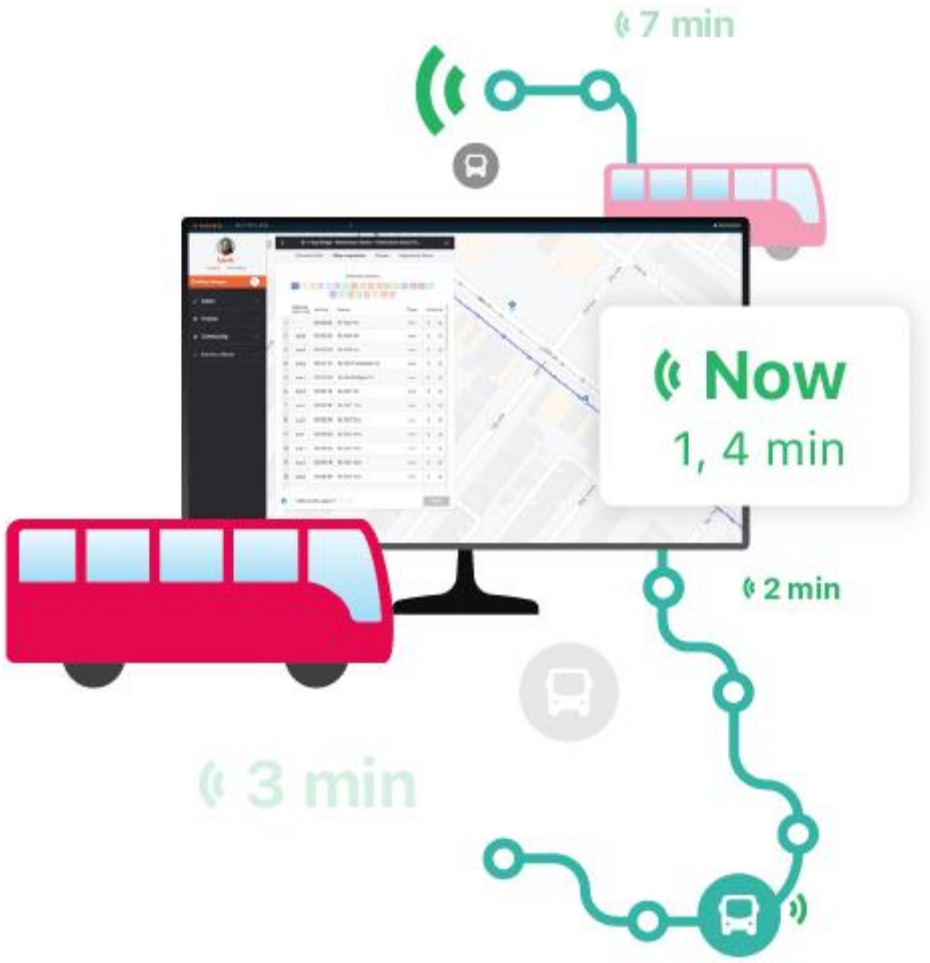
- An open standard data used for publishing public transit information to passengers.
  - Public accessibility
  - No privacy concerns
- GTFS Static data
  - Defines fixed schedules and geographic information for public transport services.
- GTFS Realtime data
  - Dynamic updates, including real-time bus speed and location.



GTFS data description

- An open standard used for publishing public transit information to passengers.
  - Public accessibility
  - No privacy concerns

- GTFS Static data
  - Defines fixed schedules and geographic information for public transport services.
- GTFS Realtime data
  - Dynamic updates, including real-time bus speed and location.

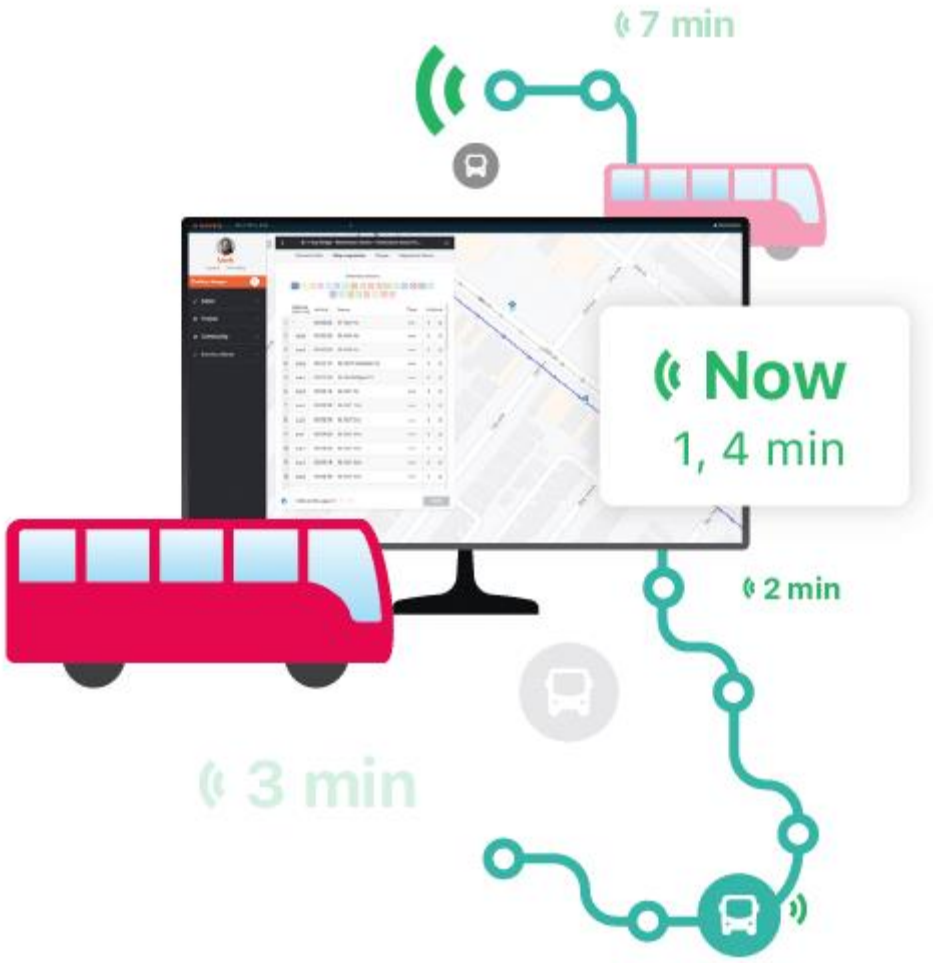


GTFS data description

- An open standard used for publishing public transit information to passengers.
  - Public accessibility
  - No privacy concerns

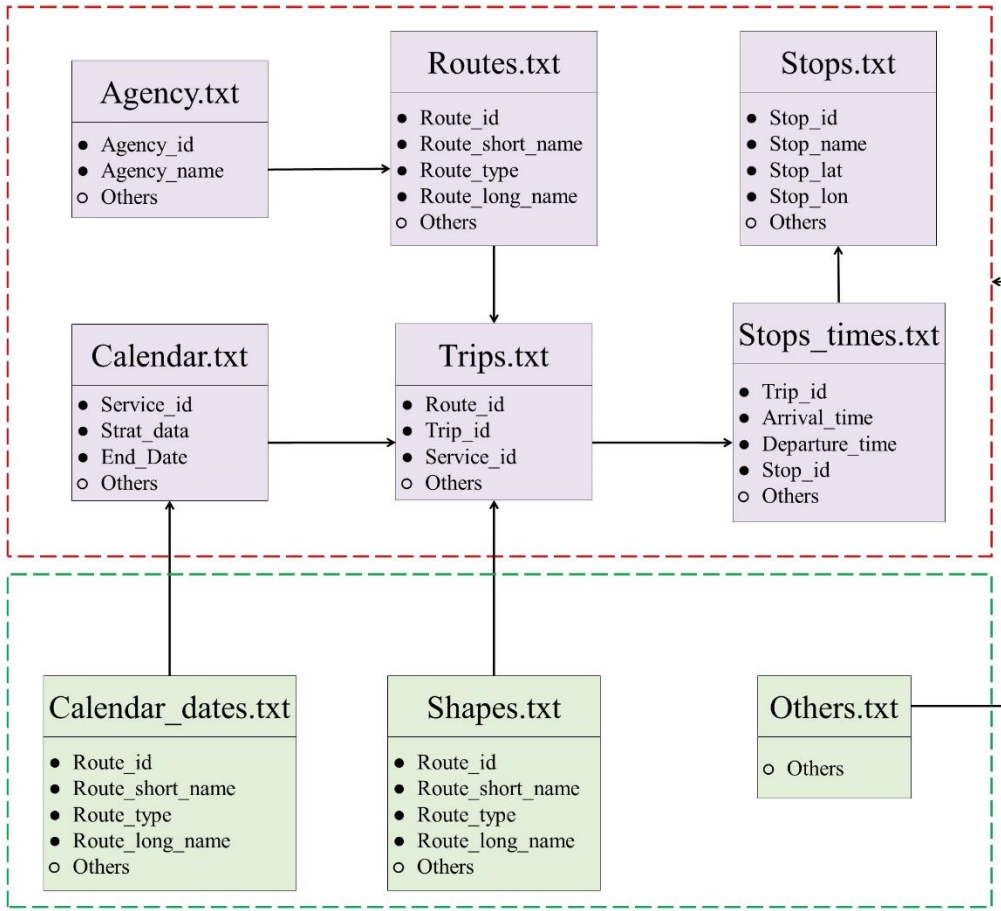
- GTFS Static data
  - Defines fixed schedules and geographic information for public transport services.

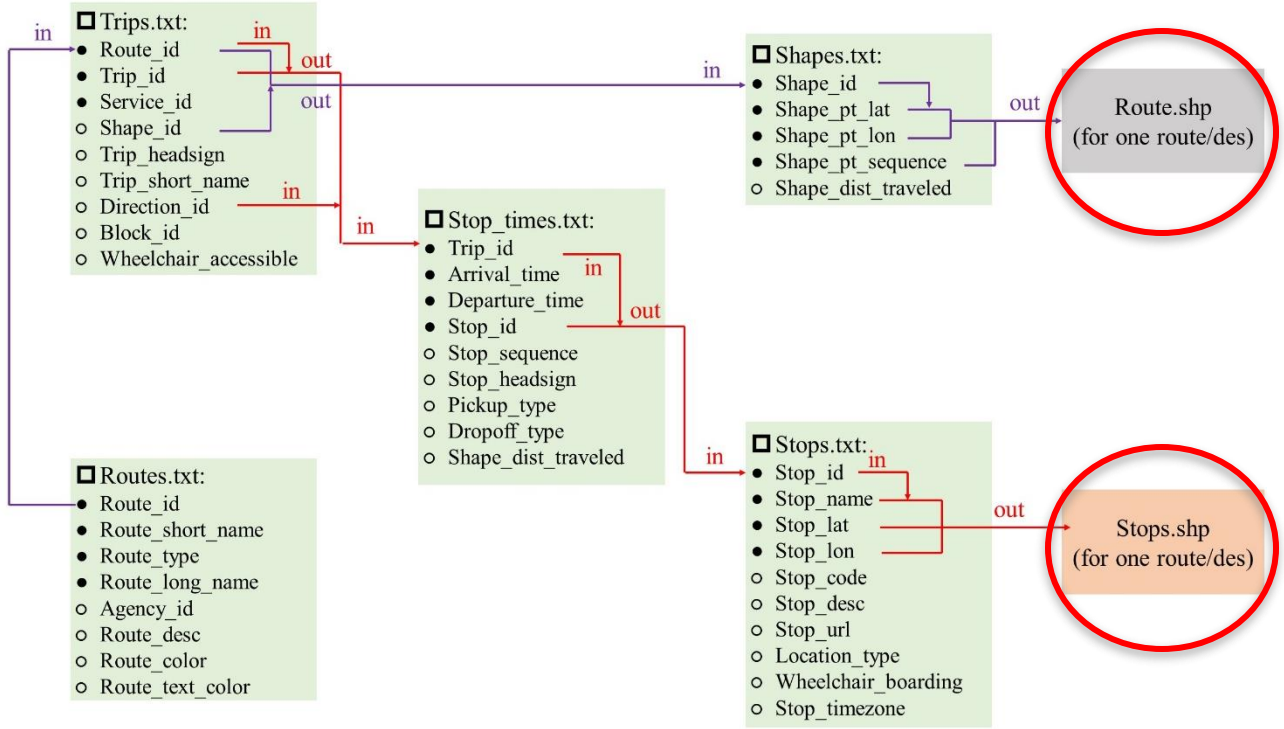
- GTFS Realtime data
  - Dynamic updates, including real-time bus speed and location.



GTFS data description

Static data





- Bus_route.shp
- Bus_stop.shp

GTFS data description

Realtime data

- GTFS Realtime
  - **Vehicle Positions**
    1. **Position: lat, lon, speed, direction, odometer**
    2. Congestion Level: Class 1~5, up to agency
    3. Occupancy Status: available space for bus, empty to unavailable
    4. Vehicle Stop Status: coming at/ stopped at / transit to
    5. Vehicle Descriptor: ID, Label, License plate
  - **Service Alerts**
    1. Time Range: time coverage for alert
    2. Entity Selector: specify exactly which parts of the network this alert affects
    3. Cause
    4. Effect: no service, detour…
  - **Trip Updates**
    1. Stop Time Update
    2. Trip Descriptor
    3. Uncertainty

Research Questions

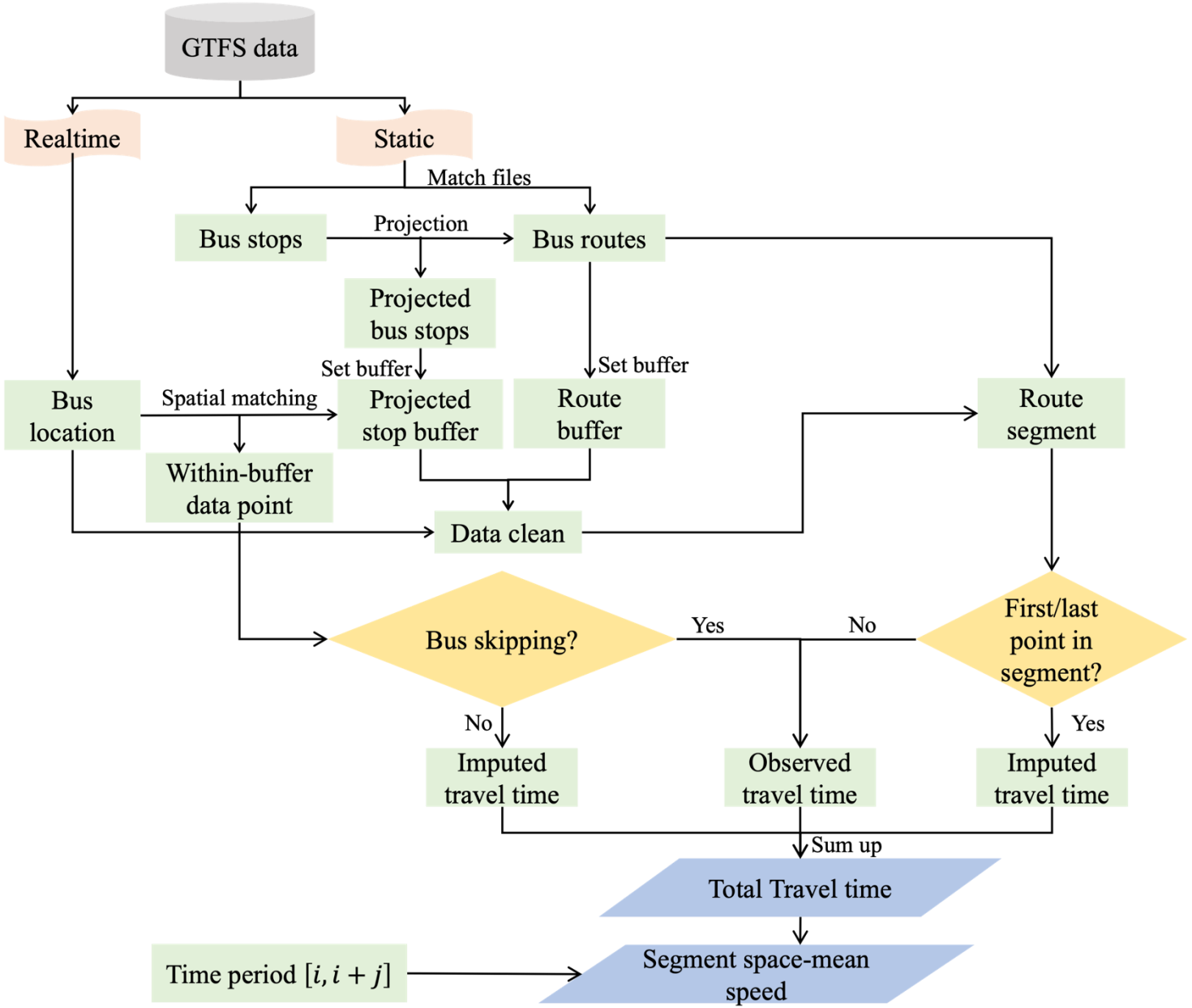
Can *GTFS Realtime Data* be used for real-time network sensing?

- Bus travel patterns → on/off boarding → different from normal traffic
- Bus travel speed → low → different speed variation pattern
- How to validate the results?

Methodology

Overview

- Metric: Speed
  - Data Collection
  - Pre-processing
  - Trip Identification
  - Travel Time Estimation



Methodology

Algorithm 1 Segment-Trip Extraction Algorithm

input GTFS Realtime data D, intersection locations P_{intxn}, bus route R, segment buffer radius b_{seg}
$S \leftarrow \emptyset$
for p_k in P_{intxn} **do**
    $s_k \leftarrow$ route segment between p_k and p_{k+1}
    $S \leftarrow S \cup \{s_k\}$
end for
$n \leftarrow$ number of segments in S
$TD \leftarrow \{TD_1, TD_2, \ldots, TD_n\}$ where $TD_1, TD_2, \ldots, TD_n$ are $\emptyset$
for s_k in S **do**
    for D_j in D **do**
        $tid \leftarrow$ trip id of D_j
        $date \leftarrow$ date of D_j
        $l \leftarrow$ bus location of D_j
        $d_k \leftarrow$ distance between s_k and l
        if $d_k < b_{seg}$ **then**
            $TD_k \leftarrow TD_k \cup \{(tid, date)\}$
        end if
    end for
end for
output TD, S

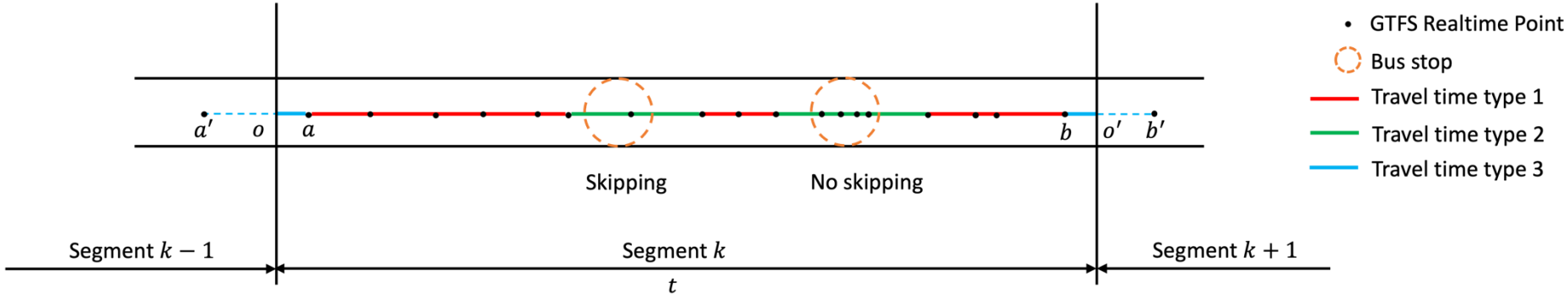
Algorithm 2 Segment Speed Estimation Algorithm

input route segments S, trips on specific segments TD, time windows TW
$n \leftarrow$ number of segments in S
$a \leftarrow$ number of time windows in TW
$V \leftarrow \{V_{11}, V_{21}, \ldots, V_{a1}, \ldots, V_{an}\}$, where $V_{11}, V_{21}, \ldots, V_{a1}, \ldots, V_{an}$ are $\emptyset$
for s_k in S **do**
    $length \leftarrow$ length of s_k
    for td_m in TD_k **do**
        $\nabla t_m \leftarrow$ estimated travel time of trip td_m
        $v_m \leftarrow length / \nabla t_m$
        $p_s \leftarrow$ the first point of trip td_m on segment s_k
        $t_s \leftarrow$ the timestamp of the p_s
        for tw_q in TW **do**
            if $t_s \in tw_q$ **then**
                $V_{qk} \leftarrow V_{qk} \cup \{v_m\}$
            end if
        end for
    end for
end for
$Speed \leftarrow \{u_{11}, u_{21}, \ldots, u_{a1}, \ldots, u_{an}\}$, where $u_{11}, u_{21}, \ldots, u_{a1}, \ldots, u_{an}$ is 0
for q in $1 : a$ **do**
    for k in $1 : n$ **do**
        $u_{qk} \leftarrow mean(V_{qk})$
    end for
end for
output $Speed$

Focus on the specific road segment → Travel speed estimation

Methodology

Travel Time Estimation

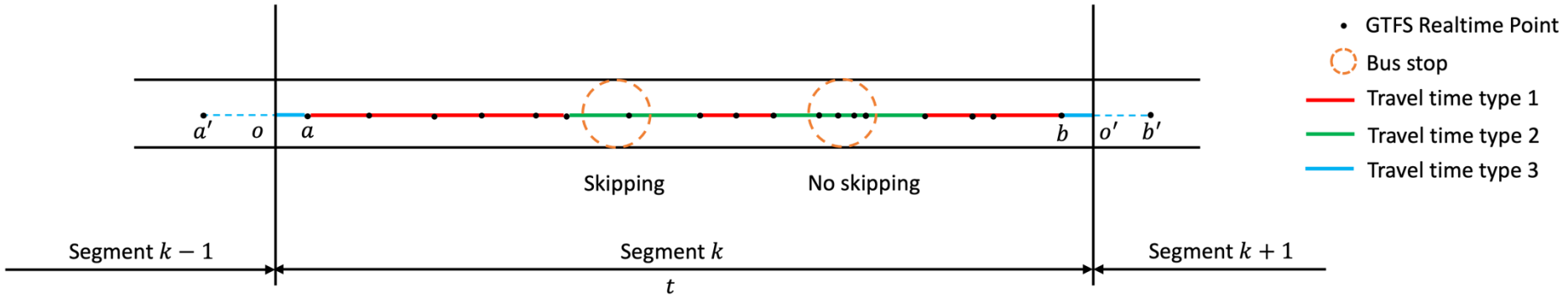


$$\Delta t_{1,k} = \sum_{i=1}^{S} \left(t_e^i - t_s^i\right)$$

Notes:
S: segment number;
P: the number of stops that the bus skipped;
M: the number of stops that the bus stopped;
s: the first GTFS Realtime data point in each subsegment;
e: the last GTFS Realtime data point in each subsegment.

Methodology

Travel Time Estimation



$$\Delta t_{2,k} = \sum_{j=1}^{P}\left(t_{e+1}^{j} - t_{s-1}^{j}\right) + \sum_{l=1}^{M}\frac{2\left(x_{e+1}^{l} - x_{s-1}^{l}\right)}{\left(v_{e+1}^{l} + v_{s-1}^{l}\right)}$$

Bus skipping

Bus stop

if $2\left(x_{e+1}^{l} - x_{s-1}^{l}\right)/\left(v_{e+1}^{l} + v_{s-1}^{l}\right) \ll t_{e+1}^{l} - t_{s-1}^{l}$

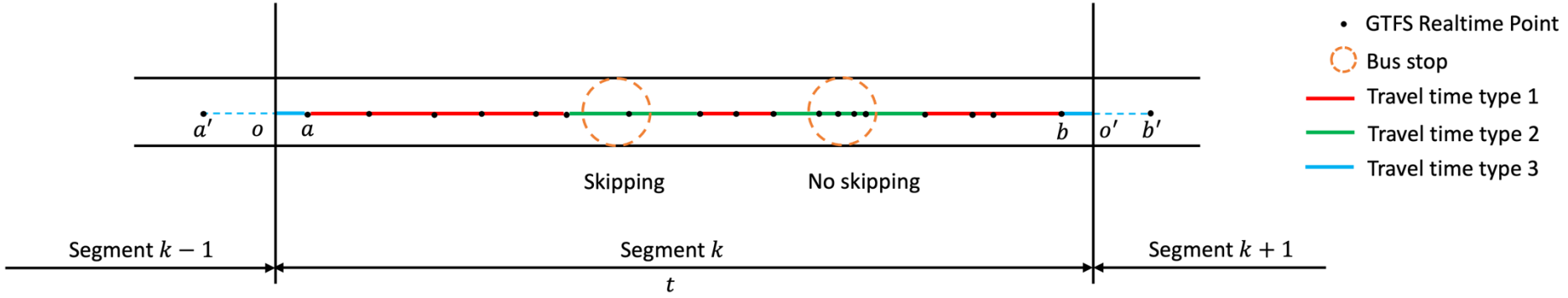
then bus stopped

15s[1]

[1] Prathibaa, K. and K. Gunasekaran, Planning of Bus Stops for Safe and Effecient Passenger Boarding and Alighting. *Planning*, Vol. 5, No. 06, 2016.

Methodology

Travel Time Estimation



$$\Delta t_{3,k} = \left(t_s^k - t_e^{k-1}\right)\frac{\left(x_s^k - x_o^{k-1}\right)}{\left(x_s^k - x_e^{k-1}\right)} + \left(t_s^{k+1} - t_e^k\right)\frac{\left(x_o^{k+1} - x_e^k\right)}{\left(x_s^{k+1} - x_e^k\right)}$$

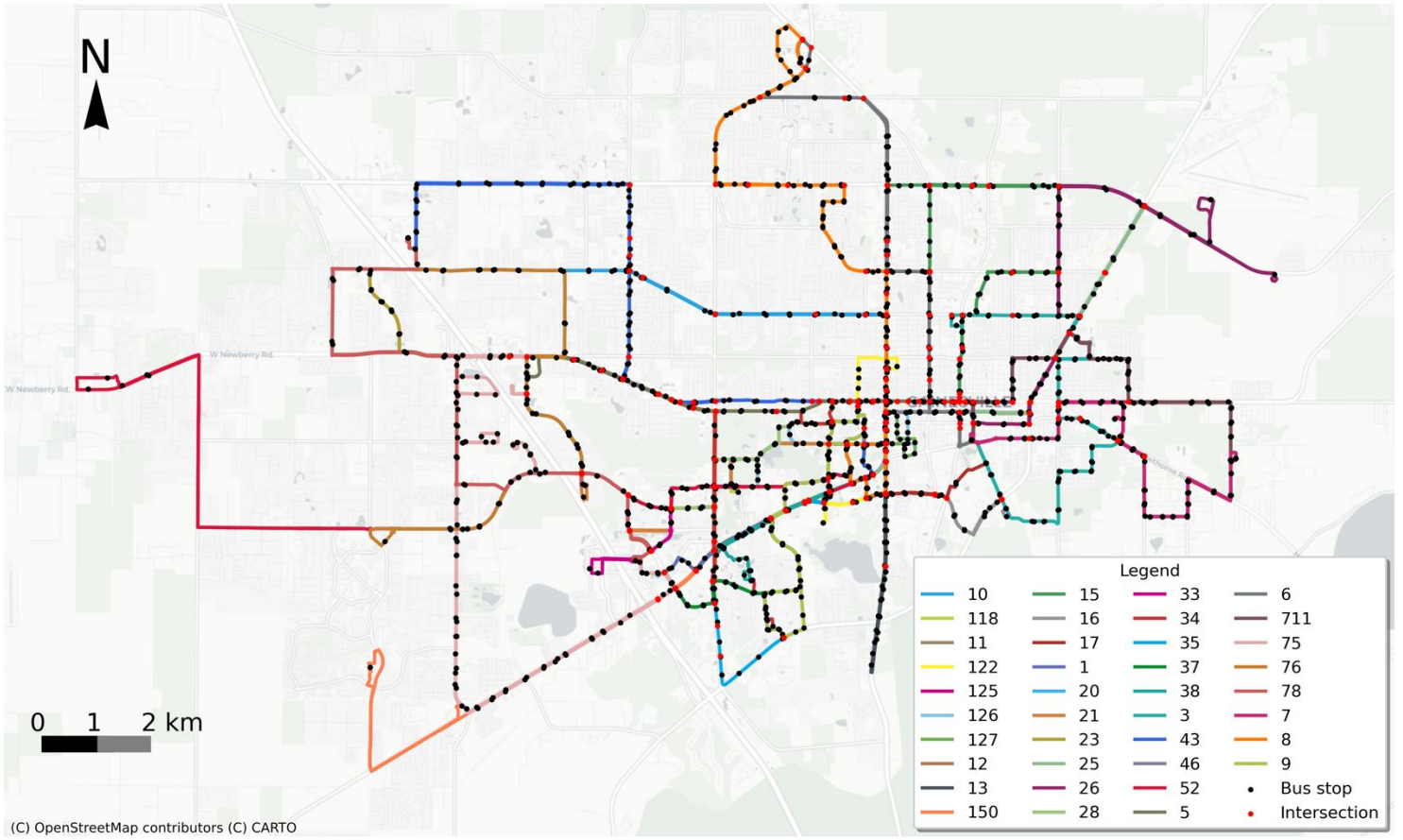
$$\Delta t_k = \Delta t_{1,k} + \Delta t_{2,k} + \Delta t_{3,k}$$

Case Study

Research Area

GTFS Realtime data:

- **Research area**: Gainesville, FL
- **Time span**: 2 weeks in Oct. 2023
  6 am. → 11 pm.
  Interval=15s
- **Records**: 1.1 M → Originally
  1.06 M → After Pre-processing



Gainesville, FL

Case Study

Validation

Data collection:

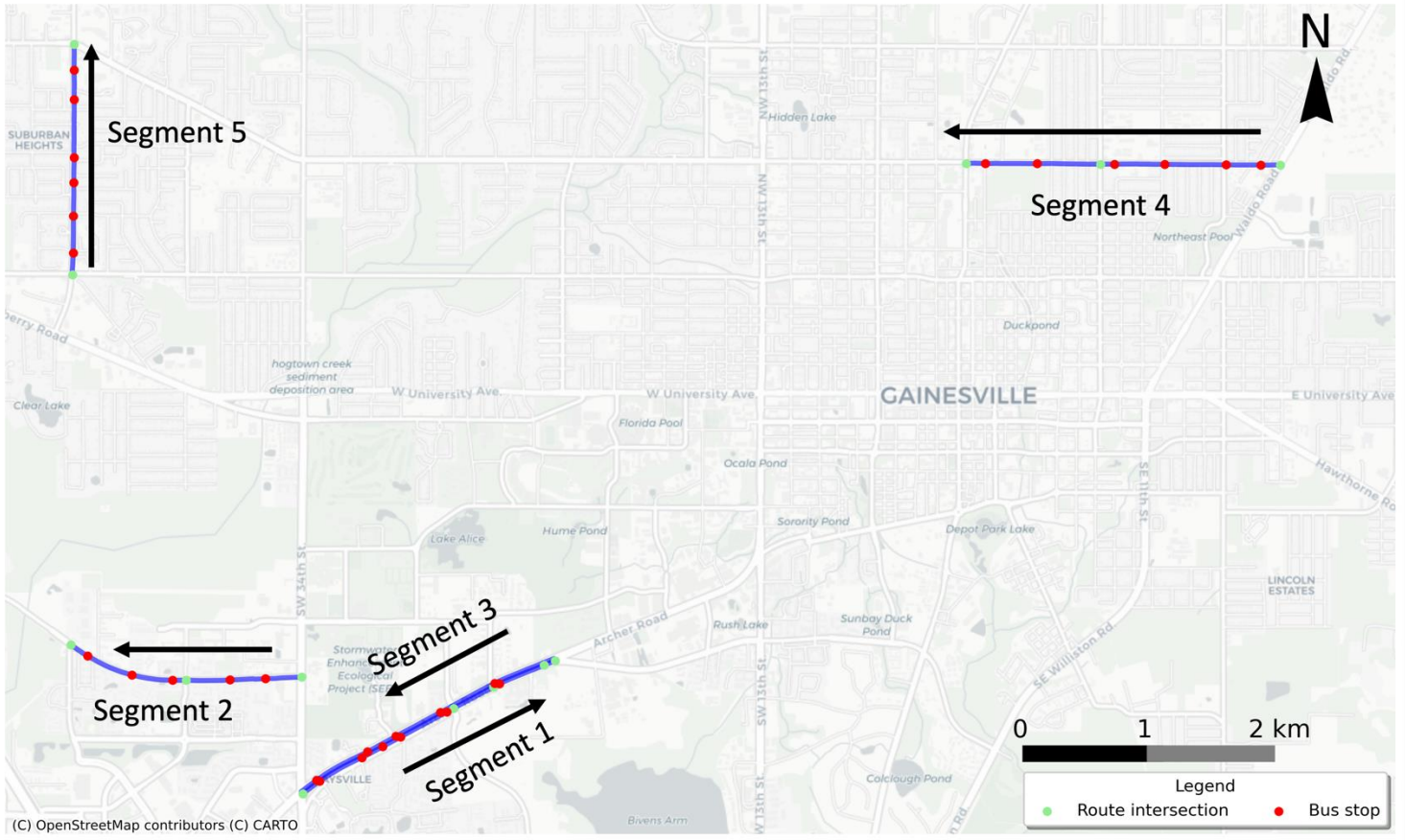
- ✓ GTFS Realtime data → Bus
- ✓ Bluetooth data
- ✓ Google Traffic data

} Car

TABLE 3: Characteristics of road segments in validation area.

Validation seg	Length (m)	Speed limit (mph)	Bus route	Road hierarchy	Data collection period
Segment 1	1997	45	1I, 38I, 150I	Arterials	10.09∼10.13
Segment 2	1673	40	20I	Collectors	10.09∼10.13
Segment 3	1987	45	1O, 38O, 150O	Arterials	10.25∼10.29
Segment 4	2190	35	3O	Local roads	10.25∼10.29
Segment 5	1606	45	43O	Collectors	10.25∼10.29

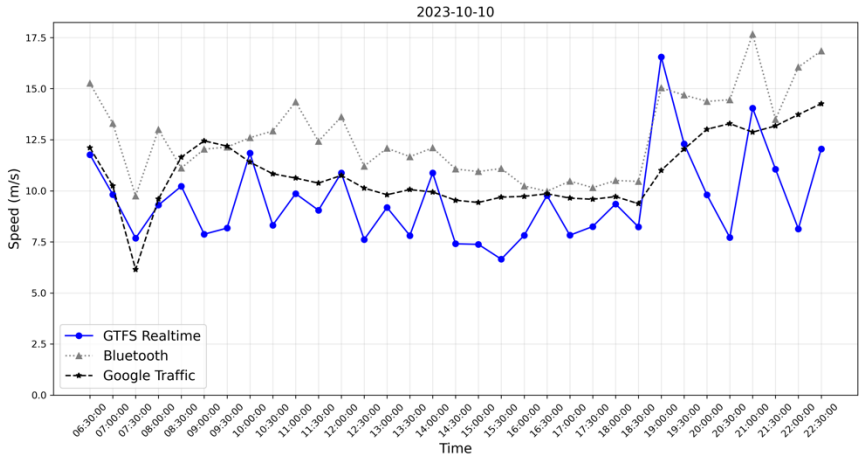
Notes: In 'Bus route' field, 'I' represents bus routes that enter the downtown direction; 'O' represents the opposite.



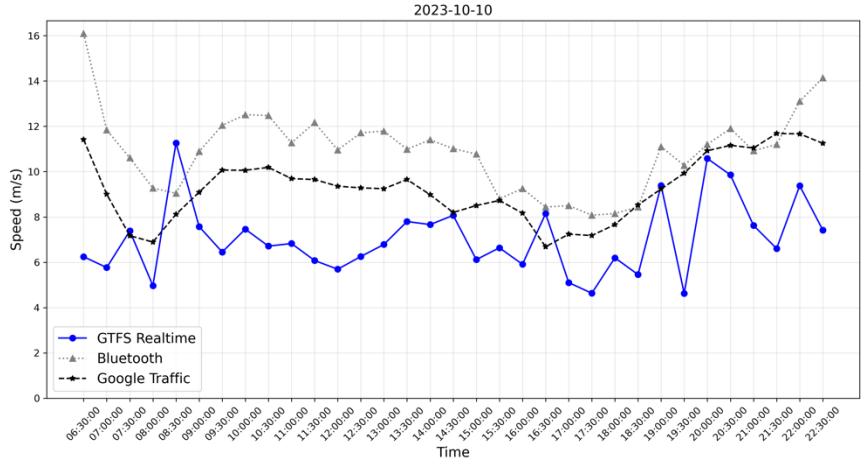
Case Study

Validation

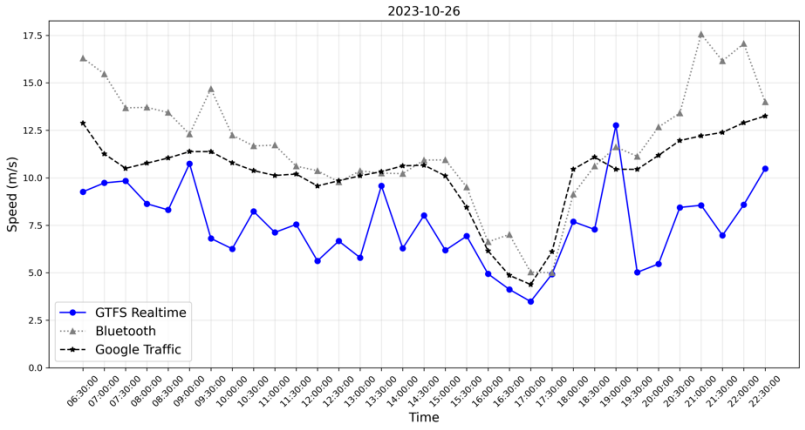
- Similar Speed variation pattern



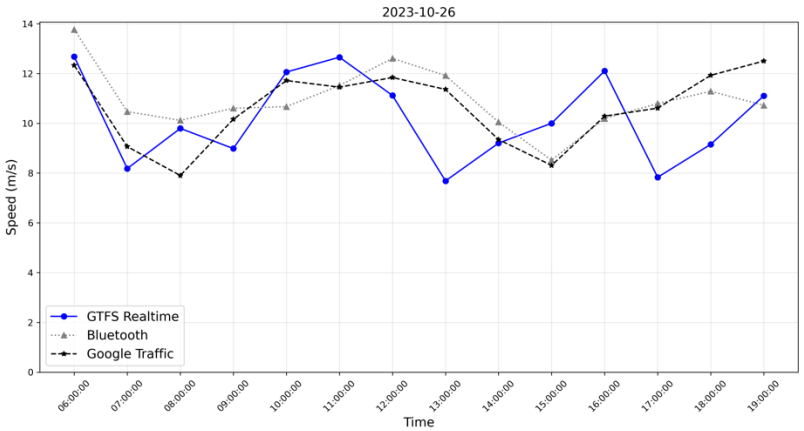
Segment 1



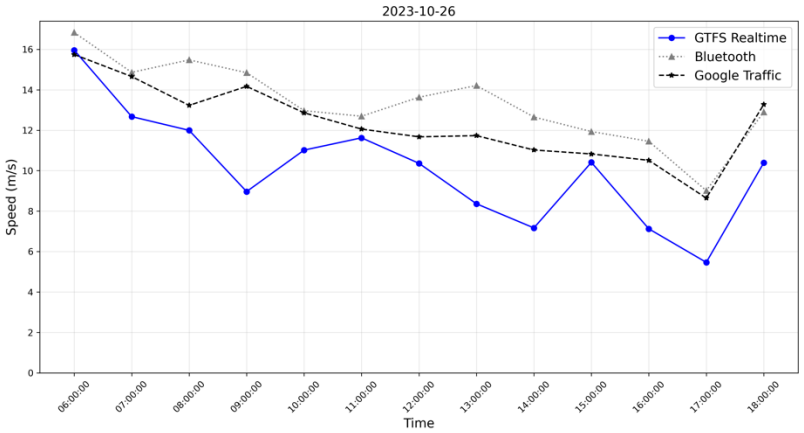
Segment 2



Segment 3



Segment 4



Segment 5

Case Study

Validation

- Hypothesis test → Kolmogorov–Smirnov (KS) test

H_0: Three datasets follow the same distribution;

H_A: Three datasets do not follow the same distribution.

$\alpha = 0.05$

Cannot reject H_0 in most of time

TABLE 4: KS-test results for GTFS Realtime data with other two datasets.

	Date	GR - BL	GR - GT
Segment 1	10/09/23	0.182(0.654)	0.182(0.654)
	10/10/23	0.273(0.173)	0.303(0.097)
	10/11/23	0.152(0.851)	0.242(0.290)
	10/12/23	0.091(0.999)	0.212(0.453)
	10/13/23	0.333(0.051)	0.303(0.097)
Segment 2	10/09/23	0.212(0.453)	0.364(0.025)
	10/10/23	0.182(0.654)	0.182(0.654)
	10/11/23	0.182(0.654)	0.182(0.654)
	10/12/23	0.273(0.173)	0.212(0.453)
	10/13/23	0.121(0.973)	0.152(0.851)
Segment 3	10/25/23	0.242(0.290)	0.393(0.011)
	10/26/23	0.121(0.973)	0.303(0.097)
	10/27/23	0.273(0.173)	0.333(0.051)
	10/28/23	0.182(0.654)	0.242(0.290)
	10/29/23	0.182(0.654)	0.242(0.290)
Segment 4	10/25/23	0.357(0.343)	0.214(0.921)
	10/26/23	0.357(0.343)	0.286(0.635)
	10/27/23	0.214(0.921)	0.143(0.999)
	10/28/23	0.143(0.999)	0.286(0.635)
	10/29/23	0.214(0.921)	0.286(0.635)
Segment 5	10/25/23	0.308(0.588)	0.385(0.300)
	10/26/23	0.154(0.999)	0.154(0.999)
	10/27/23	0.385(0.300)	0.308(0.588)
	10/28/23	0.308(0.588)	0.231(0.898)
	10/29/23	0.308(0.588)	0.308(0.588)

Notes: GR: GTFS Realtime; BL: Bluetooth; GT: Google Traffic.

Case Study

Validation

✓ Average travel time of the automobile (ATT) ~ average travel time of bus (BTT) → Linear relationship

✓ Cross validation: 80% data model fit – 20% data validation

- Correlation and Error Analysis

TABLE 5: Results of model calibration and validation.

	Correlation Coefficient (R)		RMSE		MAPE (%)	
	BL - GR	GT -GR	BL - GR	GT - GR	BL - GR	GT - GR
Segment 1	0.647	0.360	19.658	37.554	8.742	11.660
Segment 2	0.365	0.383	24.824	25.587	12.807	11.137
Segment 3	0.586	0.595	21.145	28.453	10.001	12.047
Segment 4	0.409	0.390	36.949	35.607	12.863	14.945
Segment 5	0.458	0.484	25.681	26.855	19.446	13.578

Notes: GR: GTFS Realtime; BL: Bluetooth; GT: Google Traffic.

[2] Zhou, P., S. Jiang, and M. Li, Urban Traffic Monitoring with the Help of Bus Riders. In *2015 IEEE 35th International Conference on Distributed Computing Systems*, 2015, pp. 21–30.
[3] Chakroborty, P. and S. Kikuchi, Using bus travel time data to estimate travel times on urban corridors. *Transportation Research Record*, Vol. 1870, No. 1, 2004, pp. 18–25.
[4] Kieu, L. M., A. Bhaskar, and E. Chung, Empirical modelling of the relationship between bus and car speeds on signalised urban networks. *Transportation Planning and Technol- ogy*, Vol. 38, No. 4, 2015, pp. 465–482.

Major Take-aways

- GTFS Realtime data can effectively **capture link speed variations**;
- Proposing a novel methodology to **estimate real-time traffic speed** based on GTFS Realtime data;
- Validating the possibility of extending **real-time network sensing** to the **spatial coverage** by using GTFS Realtime data.

UF Herbert Wertheim College of Engineering
UNIVERSITY of FLORIDA

POWERING THE NEW ENGINEER TO TRANSFORM THE FUTURE

Department of Civil and Coastal Engineering

Thank you !

Shangkun Jiang (jiang.shangkun@ufl.edu)

jiang.Shangkun@ufl.edu