BTE-Sim: Fast simulation environment for public transportation

Rishav Sen†, Toan Tran‡, Seyedmehdi Khaleghian‡,
Philip Pugliese⋆, Mina Sartipi‡, Himanshu Neema†, Abhishek Dubey†
†Vanderbilt University
‡ University of Tennessee at Chattanooga
⋆ Chattanooga Area Regional Transportation Authority (CARTA), Chattanooga TN, USA

***Abstract*—The public commute is essential to all urban centers and is an efficient and environment-friendly way to travel. Transit systems must become more accessible and user-friendly. Since public transit is majorly designed statically, with very few improvements coming over time, it can get stagnated, unable to update itself with changing population trends. To better understand transportation demands and make them more usable, efficient, and demographic-focused, we propose a fast, multi-layered transit simulation that primarily focuses on public transit simulation (BTE-Sim). BTE-Sim is designed based on the population demand, existing traffic conditions, and the road networks that exist in a region. The system is versatile, with the ability to run different configurations of the existing transit routes, or inculcate any new changes that may seem necessary, or even in extreme cases, new transit network design as well. In all situations, it can compare multiple transit networks and provide evaluation metrics for them. It provides detailed data on each transit vehicle, the trips it performs, its on-time performance and other necessary factors. Its highlighting feature is the considerably low computation time it requires to perform all these tasks and provide consistently reliable results.**

***Index Terms*—public transit, fast traffic simulation, model integration, data processing, road speed calibration**

I. Introduction

Whether its for a meeting or for a recreational commute, we regularly need to travel from one place to another. In doing so, the modes of transport used for every individual. In the United States, transportation accounts for 29% of all greenhouse gas emissions, among that 57% is from private vehicles, such as cars, SUVs and pickup trucks [1]. On average, public transit produces less than two-thirds of the emissions of private vehicles per passenger mile. Another issue is that public transit is not very well utilized in the United States and that there is a universal need to build and promote public transit systems to address the issues of increased greenhouse gases and emissions. Public transportation not only reduces environmental impacts but also reduces the cost of transportation for all individuals. To make it viable, the system needs to be easily accessible and tailored to the needs of the population. Transit systems are built using many individually controllable components, such as the number of buses, the routes they should run on, their times of travel, and so on, which can be difficult to visualize and configure all at once. A transit simulator is specifically designed to help this issue and by using it, we can replicate the scenario that would go on in an usually day in a city. Since we prioritize public transit, the simulator environment that we propose primarily focuses on public transportation, allowing us to assess the impact of various transportation routes in the city and how people will interact with it.

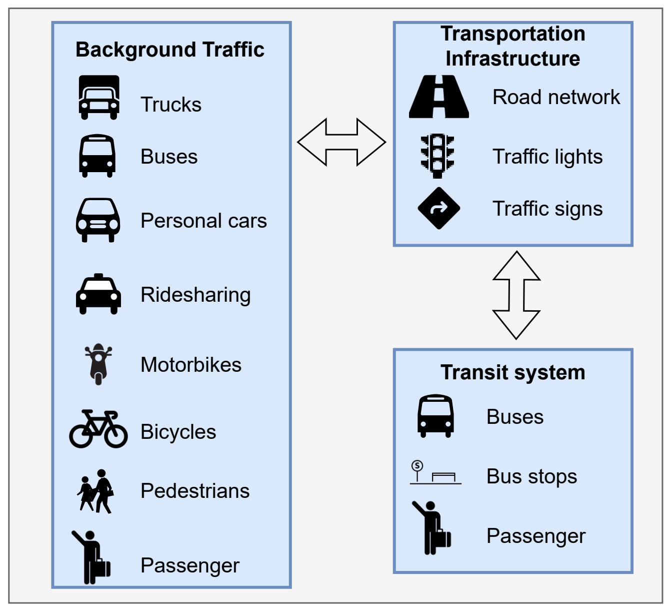


Fig. 1: Components of a transit simulation

The system relies on identifying the commuter preferences (that we are collecting through surveys) and calibrating user trips from each census block or traffic analysis zone to match the activity seen in real-life. The simulator will help to identify and address the eventual challenges of low transit efficiency by combining the complementary advantages of fixed-and dynamic-route transit services by seamlessly integrating them,

Generally, a transit simulation includes three main components: transit system, background traffic, and transportation infrastructures, illustrated in Figure 1. The transit system includes buses, bus stops, and commuters. Meanwhile, the background traffic consists of other modes of transportation, such as private vehicles, taxis, freight vehicles, and pedestrians. In addition, the transit system and the background traffic shared use the transportation infrastructure, so they affect each other.

Problem statement: We address a novel simulation scenario that only the transit system's setting changes while the rest remains stable. This problem happens in many tasks related to optimization and planning. For instance, to design optimal bus routes, optimization algorithms [11] usually execute a greedy/evolution process which repeats the simulation many times, but only the bus routes are adjusted. Furthermore, machine learning-based online routing models for on-demand transit services [12] are usually trained by a trial and error simulation process that varies only the setting of the transit system. Another example is that given the demand for a football match, planning public transportation services for this demand may require multiple times of conducting simulation evaluation [3]. Such scenarios cannot be avoided, and simulating such high-traffic scenarios is computationally too heavy for the existing system.

A naive approach is to simulate from scratch each time the transit system's setting changes. However, transit simulation is a citywide problem with a large road network and several thousands of vehicles, so it is extremely computationally expensive. The simulation in [13] took around 8 hours to simulate a day of a city-wide transit system. To improve the computing-time efficiency, we propose a module named Background Traffic Elimination (BTE) that can mimic effects of the background traffic to the transit system, and we call this method of transit simulation BTE-Sim. The proposed module can speed up the simulation 13 times while achieving competitive results in other metrics such as trip duration, bus delay, bus speeds, total distance traveled, and virtual passenger alightings.

Our approach and motivation is described in the following sections by showing how we prepare the necessary constituent datasets for the simulation, the process of simulating the system, and then we show through multiple experiments, the robustness and efficiency of BTE-Sim.

Our contributions can be summarized as the follows.

- We proposed the Background Traffic Elimination module to speed up transit simulations while obtaining competitive results compared to conventional simulations.
- We conducted comprehensive experiments to demonstrate the operation of BTE-Sim. Moreover, we presented a novel downstream task of transit simulations i.e., evaluating the OD matrices.

II. Related work

The simulator is built on the Eclipse Simulation of Urban MObility (SUMO) open source simulator [4]. It is a highly customizable, large scale simulator with capabilities to perform controls at microscopic level. It can work on city sized areas, and is highly scalable with respect to the number of components it can include. Components may include the people, cars, trucks and buses its simulating. The city's road network which includes all types of roads, drivable and non-drivable, are imported into SUMO using road data from OpenStreetMap(OSM). The roads can be custom-configured and edited further using tools from SUMO. In addition to being highly capable of integrating and finely controlling multiple parameters, SUMO also provides detailed outputs of all its components. That makes it ideal for use in our public transit simulation.

In previous literature, traffic simulators have been used to analyze transit simulations. They even extend to the use of micro-transit, or shared mobility uses [7]. While most of them are able to address the problems, they are limited in geographic area and fleet size. Urban mobility services as a whole has been studied sparsely. There have been some credible attempts to simulate city-wide transit systems as shown through Transit-Gym [13], but they are usually computationally very heavy and take a lot of processing time. Our approach tries to solve this issue by building a highly expansive, customizable, controllable, and computationally fast simulation environment focused on urban transit design.

III. Data Preparation

The simulation considers multiple inputs to set up the desired scenario:

- The road network is selected from OSM – represents the city or region's roads
- The transportation demand is obtained by generating OD pairs from publicly available datasets, like LODES (for jobs), or also from city organizations
- The bus schedules are found from GTFS
- We need to identify the number of individual cars on the road as well, to get a proper assessment of the traffic speeds

A. Road network

The urban model boundary includes Hamilton County in Tennessee, Catoosa County in Georgia, and two partial counties (Dade and Walker) in Georgia (Figure 2). The Chattanooga highway network included all interstates, other freeways, arterials, collectors, and a significant portion of the local roads. For the purpose of compiling traffic-related data, particularly the journey-to-work and place-of-work statistics, the traffic analysis zone (TAZ) is used to divide the planning region into small, relatively homogeneous areas in terms of land use and activity. TAZs are used to represent travel within a model study area because it is not practical or feasible to model individual households and employment. Housing and employment data are aggregated to the TAZ data, and the TAZs are used through the model process to calculate the origin and destination of trips in the model. To simulate travel within the Chattanooga study area, a computer network must be developed that represents the street system to be modeled. As a part of the network development process, corrections and quality checks were made to the SUMO network. Corrections made to the Chattanooga network include transit routes,traffic lights, modified disconnected intersection nodes, and repaired fragmented roadways.

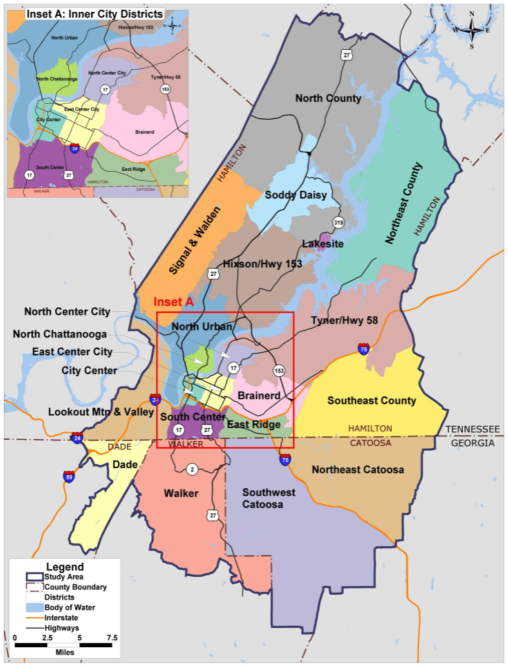


Fig. 2: Model Boundary

B. Time of day model

Four time-of-day periods are incorporated into the model stream from the destination choice to the assignment steps. The four time-of-day periods are AM peak, midday, PM peak, and off-peak. The development of the time-of-day model included identifying peak travel time periods, developing peak period factors, and developing the percentage of trips by purpose during each time period by direction. These factors were used to reflect the behavior of the traffic during the peak period. The traffic assignment step for the Chattanooga model is conducted for each time period of day.

C. Transit network development

Chattanooga Area Regional Transportation Authority (CARTA) is the main provider of public transit services in the region. CARTA operates 13 fixed-route bus routes, two dial-a-ride neighborhood routes, a free shuttle route around the University of Tennessee Chattanooga campus, and two free downtown electric shuttle routes. All fixed-route bus routes (including three shuttle service routes) were modeled in the transit network.

D. Transportation demand

Urban area travel demand models are important tools for the analysis of transportation plans, projects, and policies. Travel demand modeling practices vary significantly between Metropolitan Planning Organizations (MPOs) in the United States. In general, traditional 4-step models require less time, data, and resources to develop and validate. More advanced travel demand estimation techniques, such as the activity-based approach, usually require more resources to estimate, validate, and use, but have the benefits of improved sensitivity to policy changes. The geographic area under consideration is Chattanooga, Tennessee. The city is divided into census tracts. We find the movement matrix of people travelling for jobs (LODES) from census tract to census tract. The generated origin-destination(OD) pairs are used as requests for our solver. Each row in either dataset represents a single trip by one person. The trip represents movement to the job location and then back to home, at certain times of the day, which are sampled from a given set of regular job start and job end times.

The data are a congregation of the geographic data of the area (from OpenStreetMap), people movement, census(LODES) data, residential and work locations, and open-source building footprint data (we use Microsoft's open data). We use a combination of the above-mentioned datasets to arrive at the custom Transportation demand.

E. Schedules

Our motive here is to represent all of the city/county's area and model the movement of people to and from each specified subregion. This is achieved by discretizing space and time to create the data set. Each entry in the dataset refers to one person moving from their inital(home) location to their final(usually their workplace, may be different for different types of data sources) location. The data sources can be divided into two primary categories: (a) Data source for mass movement of people, usually on a census tract or census block group scale, (b) Data source for housing, work, and miscellaneous building locations throughout the region under consideration.

F. People's Movement data

Usually such datasets are obtained from census bureaus, which collect county-wide information on people's travel patterns and their preferred destinations. We use the LEHD Origin-Destination Employment Statistics (LODES) [15] data, which is publicly available from the United States Census Bureau. These may change over the years. This data is aggregated for each census block group, having the number of people travelling between census block groups. It also contains the number of workers, classified into age groups, wage groups, and industry sector.

Another source could be individual organizations that use tracking technologies to record the activity of their user base. These give us a temporal trend of the population's travels but are not usually spatially widespread and mostly confined to cities. One such dataset is Safegraph [8], which is also on a census block group scale of aggregation. This counts the number of people moving between their residence census block groups and different types of destination (like, offices, grocery stores, entertainment places), which is also in a specific census block group. This includes the frequency and count of visits, and hence the a robust temporal distribution.

G. Building locations

As we have seen in the previous subsection, the data is usually aggregated on the census block group scale, and we have no specific information about the exact home or work locations that are needed to formulate specific Origin-Destination pairs. It is imperative we find the locations of individual houses and workplaces in the concerned census block groups. Two such hierarchical methods are described here. OpenStreetMap [6] provides a labelled collection of buildings. They are individual geometric shapes on a two-dimensional (2D) plane. These buildings are then represented as a point, which are the centroids of the building's shape. Since these buildings are tagged, we can classify them into homes or workplaces and form a primary notion of the exact locations of where people are moving. One drawback of this method of data collection is that, for smaller or less populous areas, the buildings are not properly tagged, making the dataset very small and restricted to major population centers.

Census block groups that lie on the outskirts or far from cities are usually missing any sort of tagged buildings, hence a distinct lack of home or residential buildings can be seen. We are unsure of the exact locations of where people are starting and ending their trips. To get some spatial clarity and increase the granularity of the people's movement locations, we use a secondary layer of building location information. This data layer usually covers a large number of buildings in both urban and rural areas but is not tagged, that is, buildings cannot be specifically classified as homes, workplaces, or if they are used for other purposes. Microsoft produced a dataset for vector creation throughout the US in 2018 [14] that was generated from aerial images available to Bing Maps using deep learning methods for object classification [5]. We use the US Buildings Footprint to find the set of untagged buildings. Combining the tagged and untagged buildings, we can essentially cover the landscape under consideration.

H. Calibrated traffic speeds

A microscopic model describes the movements of specific combinations of vehicle and driver. These behaviors are the outcome of the features of drivers and their vehicles, the interactions between drivers, driver-road interaction and road characteristics, external factors, and traffic rules and control. This necessitates calibrating the parameter values so that the simulator may be used for simulation analysis in the desired context. Finding the optimal parameter values that minimize the discrepancy between the simulated and actual output values is, in general, the calibration procedure. Because of the impact of the parameters related to driver and vehicle on speed, a focus is put on an accurate portrayal of velocities in the simulation. A SUMO simulation of Chattanooga is compared to real-world vehicle speed data from INRIX as the foundation for evaluating the simulated behavior. The parameters of a transportation simulation model must undergo a meticulous calibration procedure to ensure that the model's output is as accurate as feasible. Figure 3 describes the workflow of the calibration algorithm. The calibration charts are presented below.

Traffic demand is the description of the vehicles that will circulate on the simulated map. The O/D matrices for this study are provided by the Chattanooga Hamilton County regional planning agency. An O/D matrix provides traffic flows (often vehicle flows) from each origin to each destination. With O/D matrices, traffic may take several paths to complete the trip from the origin to the destination. As there are 909 TAZ in our study area, we have 909-by-909 O/D matrices. Using demand generation tools of the SUMO DUArouter, OD2trips, the O/D matrices were imported and split into single vehicle trips. The simulation needs to figure out how to get from the origin edge to the destination edge for a collection of vehicles with a set of origin-destination relations (trips). In a network with high traffic, the difficulty of selecting optimal routes that account for journey times is known as user assignment. To address this issue, SUMO offers a variety of options. Using duaIterate to compute a user equilibrium, that is, it tries to identify a route for each vehicle such that no vehicle may lower its trip cost (typically the travel time) by taking a different route. It accomplishes this iteratively by: 1- using duarouter to route automobiles in a network with the lowest known edge costs (starting with empty-network travel times). 2- Invoking sumo to mimic "actual" travel times based on the computed routes. The resulting edge costs are utilized in the net routing stage. These repetitive procedures are known as Dynamic User Assignment (DUA) and are often used in combination with a traffic simulation framework. Furthermore, a specified network was created in SUMO, and detectors were installed to collect output such as speed. The detector computes the values by calculating the entry and exit timings of the vehicle. Next, a sensitivity analysis was conducted to determine the relevant factors that can have a significant impact on particular results. Then, a genetic algorithm (GA) model was created to determine the optimal values for every relevant parameter. The primary distinction between GA and other conventional search algorithms is how the computer selects sites. Classic search algorithms select points more arbitrarily and iterate until specific requirements are met, whereas GA selects points randomly and then mutates them to produce a new set of values [9]. In GA, a random parent from the original population would be picked and assigned a fitness value according to the fitness function. This parent undergoes mutation and cross-over to produce the child whose fitness value is determined and compared to the parent. If the fitness value of the child exceeds that of the parent, the child will become the parent and the cycle will continue. If lower, the parent endures more mutations until its fitness value falls below that of the child. As mentioned,the performance of the calibrated model was evaluated using speed as the effectiveness metric. For this purpose, speeds derived from the calibrated SUMO model and the INRIX dataset were compared and the errors were quantified in terms of Root Mean Square Error (RMSE). Finally, the calibrated model is used as background traffic.

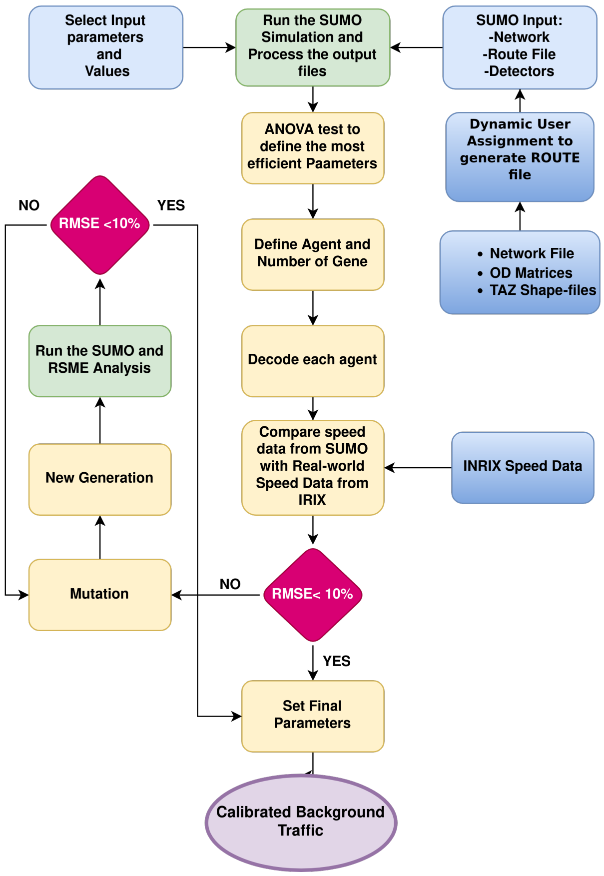


Fig. 3: Calibration procedure

IV. Simulation

Public transit simulation is done conventionally by feeding all data at once. This makes it slower to complete and requires more computational processing. To make this task easier and much faster, we leverage the use of calibrated background speeds in our proposed system, simulation with background traffic elimination (BTE-Sim). Both methods are shown here.

A. Conventional method - simulation with background traffic

This method of simulation is quite straight-forward, in the sense that it runs all the collected data about transit vehicles, personal vehicles and the people moving all at once to simulate a city's scenario, and is based on [13]. As we have shown earlier, we collect data of different types to run the system. These are fed into the python intermediate which structures and converts them into SUMO (an open source transit simulator) usable formats as shown in Fig.4. SUMO performs the simulation for the desired periods (usually 24 hours) and outputs the movement details of each bus – like their speed, acceleration, distance covered, and so on. We can then use this data to find the average speed of each traffic lane and calibrate those lanes to the found speeds. This is the speed due to the background traffic (all other vehicles, except the buses). This is used further down the line in the updated type of simulation, which is simulation using background traffic elimination.

B. Simulation with Background Traffic Elimination (BTE-Sim)

This form of simulation requires the same inputs as previously used, along with the calibrated background speeds, as shown in Fig.5. This background traffic speed can be received from one of three sources:

a) the background traffic speeds of the simulation discussed above
b) the automated passenger count (APC) dataset for the buses of the transit agency
c) INRIX, which is a dataset containing the speeds of vehicles on discrete road segments at very fine time intervals.

We can choose the background speed from any one of the above sources and we show in Experiment 1 [sec.V-A], amongst the choices, which background speed data is preferred for our system. Since we do not need to simulate the movement of all vehicles, and focus only on bus movements, it can run in a much shorter duration. Thus, greatly reducing the execution time of a day's simulation, providing faster results (around 5 minutes on a regular PC). It functions similar to the previous scenario, but the lane speeds are individually modified here, using a component of SUMO, called the TraCI (traffic control interface) client. It allows "on-line", as in while the sim is running, to control the lanes and buses. The outputs are details of the movements of the bus, which can be used to check for optimality or for other purposes, such as energy calculations [10].

V. Experiments and Results

Here, we discuss the multiple scenarios that we put the BTE-Sim system through. We aim to show that our system performance is comparable to the current standards (based on Transit-Gym [13]) in terms of results and its time of completion is significantly faster than other methods. For this, we test the system on data collected from Chattanooga, Tennessee, USA. We have information about its bus routes, its traffic data through INRIX, automated passenger count (APC) data for the time of arrival at every bus stop, the origin-destination(OD) travel data from (a) the city's planning agency, and (b) synthetically generated OD datasets from census data. We put together these data into four different experimental scenarios and show the BTE-Sim performs better than the existing system. All the experiments for both Transit-Gym and BTE-Sim are performed using the same computing power. We use the absolute error on Time of Arrival (ToA) at bus stops, which shows how early or late a simulated bus arrives at a bus stop compared to its designated time. This measure shows the relative accuracy and reliability of the simulation. Since BTE-Sim requries prior information of the The road segment (edge) speed, in our first experiment we determine which edge speed data is more suited to run the simulations.

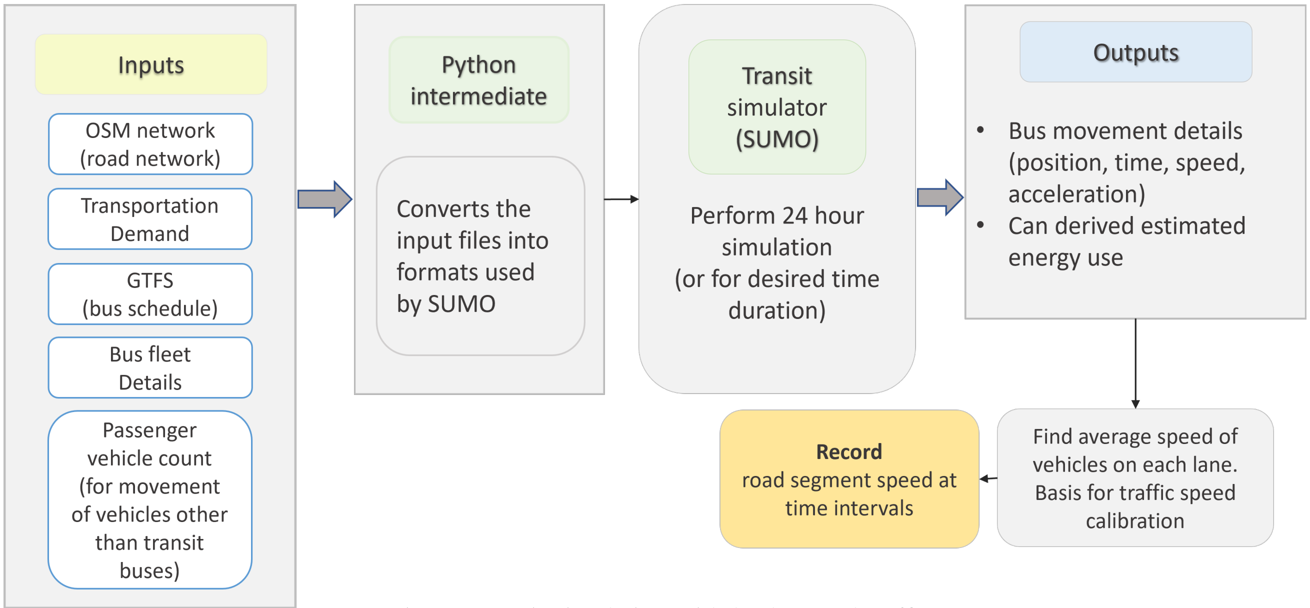


Fig. 4: Transit simulation with background traffic

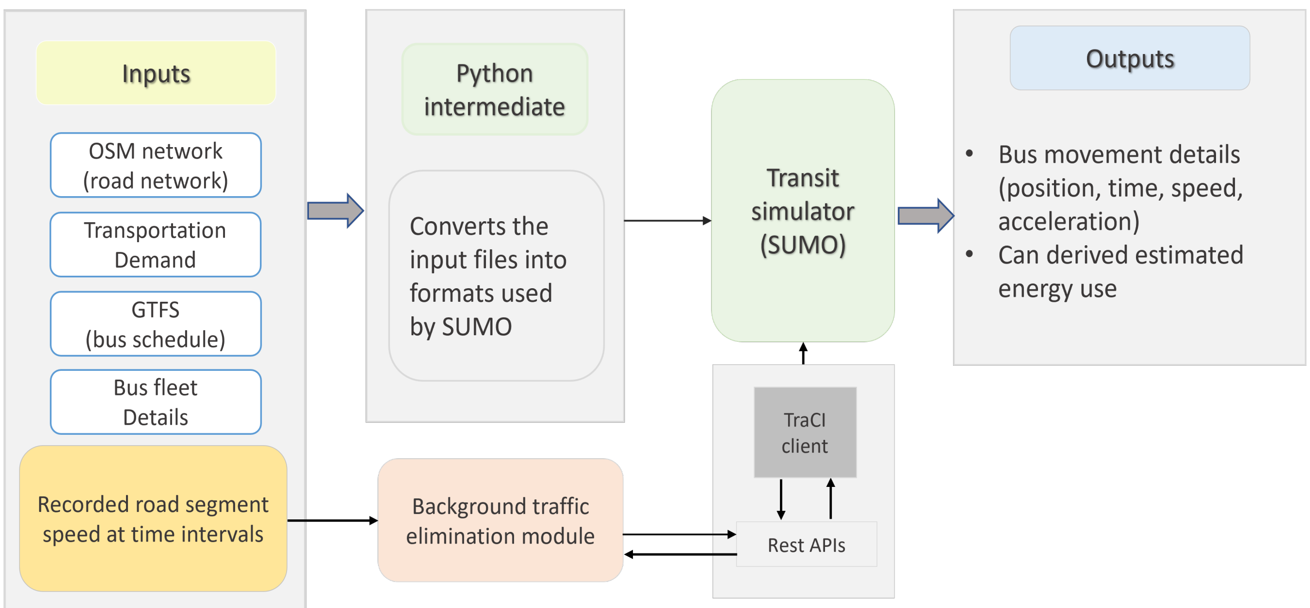


Fig. 5: Transit simulation without background traffic

A. Experiment 1: BTE-Sim with different sources for edge speed data

Setting: As mentioned above, the edge speed data of BTE-Sim can be collected from different sources. In this experiment, we use BTE-Sim from three sources to simulate the transit system on January 11, 2022. The first source is from a prior run of Transit-Gym which uses the OD matrix to generate the background traffic. The second source is the INRIX traffic and road speed service [2] which is collected from connected cars and mobile devices, cameras and sensors on roads, and major events expected to affect traffic and is available through INRIX IQ, a SaaS-based cloud platform. The third is APC, which as described earlier, records the time that buses arrive at each bus stop. In all the cases, we know the travel time of buses between two bus stops from the data. In addition, we also know the distance of the bus stops, so we can estimate the average speed of the buses when moving from any bus stop to another.

Result: In Fig. 6, the absolute ToA error is amplified for APC data, while INRIX provides comparatively better results. We could have chosen the other, independent data sources as well, but BTE-Sim performs the best when using the background traffic times generated using Transit-Gym (marginally better compared to INRIX data). Thus, we choose the background traffic speeds from Transit-Gym to be our traffic speed data for the rest of the simulations. It should be noted that we can choose to use the other, independently available data sources, like the INRIX data in our case, as the background speed. This would remove the dependence on the background speeds from Transit-Gym.

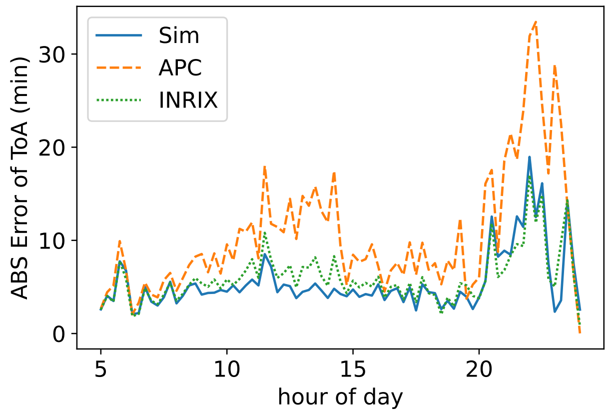


Fig. 6: Comparing background traffic sources for BTE-Sim

B. Experiment 2: BTE-Sim achieves competitive results compared to Transit-Gym

Setting: We conducted simulations for the city of Chattanooga and for the day of January 11, 2022. We used real data from the day's transit operation in the simulation. To generate the background traffic for Transit-Gym we use the OD travel matrix provided by the city's planning agency.

Result: To evaluate the simulation, we focus on the difference of Time of Arrival (ToA) at bus stops between the simulation and real world times, on the same date. It is measured in minutes. Fig. 7 shows the absolute error of ToA of Transit-gym and BTE. This shows that, using historical edge speeds, BTE can mimic the effects of background traffic, without which it would have no delays. But in accommodating background traffic, it has the capacity to change dynamically with the traffic conditions that may arise in the city. From this we can also infer that BTE performs comparatively better at simulating the movement of buses across the city, having a consistently lower ToA than Transit-Gym for an entire day's operation. In addition to the ToA performance, BTE-Sim is able to provide an extensive analysis of transit performance. Fig. 8 presents the performance of the transit system in Chattanooga on January 11, 2022 by BTE-Sim.

C. Experiment 3: BTE-Sim improves the simulation time

Setting: For the city of Chattanooga, we investigate scenarios of varying number of vehicles on the road. There could be more vehicles, and to mimic that we increase the number of vehicles in the OD matrix.

Result: Table I records the execution times of both simulation methods. For a baseline of 100,000 vehicles, we can notice that BTE-Sim runs more than 12 times faster than the traditional Transit-Gym. With a varying number of vehicles, the computation speed of the simulations also changes drastically.

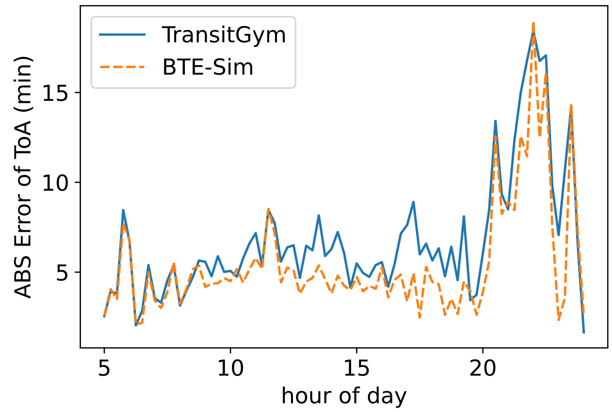
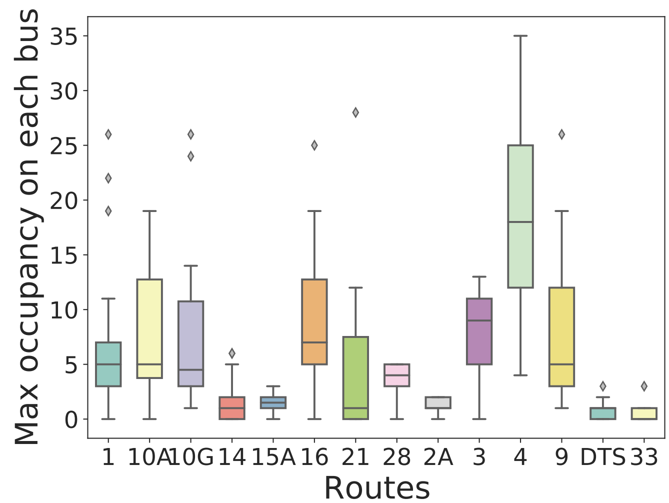
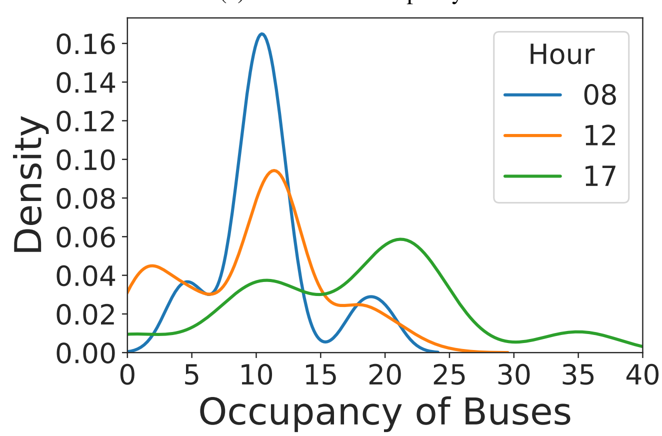


Fig. 7: Comparing Transit-gym and BTE-Sim on absolute error of Time of Arrival



(a) Maximum occupancy



(b) Density of Route 4

Fig. 8: Analysis examples for the transit system of Chattanooga on January 11, 2022, using BTE-Sim

As the number of vehicles increased by a factor of 4, Transit-Gym execution time increased by 8 times, whereas BTE-Sim computation time increased by a factor of 2. BTE-Sim runtime increases minimally with a huge increase in vehicular traffic, and is highly adaptable to traffic volume changes and can be re-run for increased traffic scenarios without much time penalty.

#Vehicles	Transit-Gym	BTE-Sim
100K	27.7 minutes	2.21 minutes
400K	4 hours 4 minutes	5.11 minutes
800K	16 hours 51 minutes	7.81 minutes
1400K	41 hours 18 minutes	8.27 minutes

TABLE I: Simulation time of Transit-Gym and BTE-Sim for scenarios with different number of vehicles

D. Experiment 4: Transit simulation can evaluate OD matrices

Setting: The demand for transit (that is, the number of transit commuters and their travel routes) can change between dates. We use three different OD datasets to simulate the scenario of changing OD demand data. The datasets are disparate as each of them contain a varying number of people moving, to different locations throughout a day. For the same day as January 11, 2022 in Chattanooga, we tested these OD variations to show the absolute ToA error.

Result: BTE-Sim shows minimal differences in the three OD situations as seen in Fig. 9 The changes in absolute ToA error are minuscule for a day's operation. This shows that BTE-Sim is well equipped to maintain steady simulations even in varying situations.

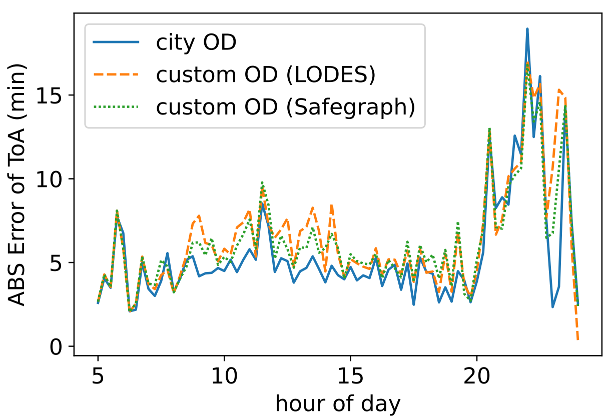


Fig. 9: BTE-Sim for different OD data

E. Experiment 5: BTE-Sim simulates different dates

Setting: We demonstrate the operation of BTE-Sim by simulating for a week from Jan. 10, 2002, to Jan. 16, 2022. Note that the transit setting for various dates is different. For example, a trip may be offered on Monday but unavailable on Tuesday.

Result: As we can see in Fig 10, the simulated values are usually very less scattered, with the ToA values having very short inter-quartile ranges for a given day, with a comparatively higher dispersion on 01-12. Both the mean and maximum values of the absolute error of ToA are consistently under 10 minutes. With these we can confirm that BTE-Sim has a very low error margin in simulating regular traffic and the operations of the transit systems on different days.

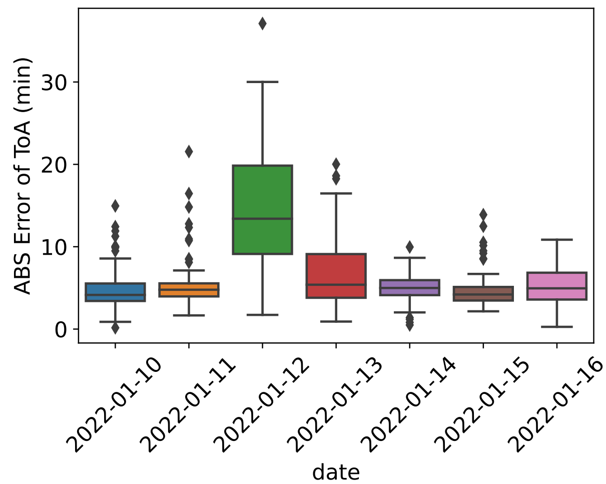


Fig. 10: BTE-Sim over different dates

Through all the experimentation we successfully show that BTE-Sim is capable of performing consistently under dynamic scenarios for city wide transportation operations. It is able to handle large scale traffic, and people movements along with its primary focus of public transportation. In comparison to Transit-Gym, we can see that the BTE-Sim system performs far better, as it is intentionally focused on catering the needs of the transit systems only. The simulation time is significantly faster than existing methods and continues to be so under increased traffic load.

VI. Conclusion

Through the experiments conducted, we demonstrate the ability to simulate a region's transit system while taking into account its population and non-transit users. The method demonstrated gives us fast, reliable and temporally and spatially accurate results about public transportation. We can use the data obtained here to improve existing systems, increasing efficiency while serving more passengers. We can support the cause of pivoting away from private vehicles and making public transit more accessible. It can also be used for purposes of proposing new transit routes, changing sections of existing routes, and estimating energy consumption on each trip. New types of buses and modes of transportation can be tried out in the process.

VII. Acknowledgments

This material is based upon work supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), under Award Number DE-EE0009212 and National Science Foundation under Award Numbers 1952011 and 1818901.

References

[1] Tina Hodges. *Public transportation's role in responding to climate change*. Diane Publishing, 2010.

[2] INRIX. Inrix delivers products for the automotive and transportation industries such as real-time parking and traffic information and solutions that facilitate the safe testing and deployment of autonomous vehicles. https://inrix.com/, 2021.

[3] Pramesh Kumar and Alireza Khani. Evaluating special event transit demand: A robust principal component analysis approach. *IEEE Transactions on Intelligent Transportation Systems*, 22(12):7370–7382, 2021.

[4] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wiessner. Microscopic traffic simulation using sumo. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2575–2582, 2018.

[5] Microsoft. Us building footprints, 2018.

[6] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org, 2017.

[7] Luca Quadrifoglio, Maged M Dessouky, and Fernando Ordóñez. A simulation study of demand responsive transit system design. *Transportation Research Part A: Policy and Practice*, 42(4):718–737, 2008.

[8] SafeGraph. Places schema. retrieved from https://docs.safegraph.com/docs/places-schema, 2020.

[9] Yadavilli Sashank, Nitin A Navali, Arjuna Bhanuprakash, B Anil Kumar, and Lelitha Vanajakshi. Calibration of sumo for indian heterogeneous traffic conditions. In *Recent Advances in Traffic Engineering*, pages 199–214. Springer, 2020.

[10] Rishav Sen, Alok Kumar Bharati, Seyedmehdi Khaleghian, Malini Ghosal, Michael Wilbur, Toan Tran, Philip Pugliese, Mina Sartipi, Himanshu Neema, and Abhishek Dubey. E-transit-bench: simulation platform for analyzing electric public transit bus fleet operations. In *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems (e-Energy 2022)*, 2022.

[11] Amutheezan Sivagnanam, Afiya Ayman, Michael Wilbur, Philip Pugliese, A. Dubey, and Aron Laszka. Minimizing energy use of mixed-fleet public transit for fixed-route service. In *AAAI*, 2021.

[12] Amutheezan Sivagnanam, Salah Uddin Kadir, Ayan Mukhopadhyay, Philip Pugliese, Abhishek Dubey, Samitha Samaranayake, and Aron Laszka. Offline vehicle routing problem with online bookings: A novel problem formulation with applications to paratransit. In *IJCAI*, 2022.

[13] Ruixiao Sun, Rongze Gui, Himanshu Neema, Yuche Chen, Juliette Ugirumurera, Joseph Severino, Philip Pugliese, Aron Laszka, and Abhishek Dubey. Transit-gym: A simulation and evaluation engine for analysis of bus transit systems. In *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 69–76. IEEE, 2021.

[14] Bing Maps Team. Computer generated building footprints for the united states, 2018.

[15] United States Census Bureau. Longitudinal Employer-Household Dynamics - Origin-Destination Employment Statistics. https://lehd.ces.census.gov/data/, 2017.