Does real-time transit information save time? Analyzing the impacts of real-time transit apps on users waiting time

Luyu Liu

The Ohio State University

**Abstract:**

Real-time information (RTI) and Real-time transit apps (RTA) have been proven to have significant impact on passengers’ waiting time and user experience. Although precious research thoroughly surveyed and simulated the overall impact, few studies investigate the impact’s geographic and temporal pattern. In this paper, we utilize General Transit Feed Specification (GTFS) real-time data for the buses’ high-resolution real-time status. We first introduce several trip planning strategies (TPSs) to simulate RTA and non-RTA users’ incidence behavior and the concept of *volunteered optimization*. Then, we optimize *prudent relaxation* strategy based on RTI and we calculate the difference between different TPSs to measure RTA users’ waiting time reduction. Moreover, we visualize and analyze the waiting time difference and find the universal presence of marginalized stops near the originating stops. The results prove that RTI and RTA can significantly decrease waiting time for users and it demonstrates great variation geographically and temporally. The methodology also shows great potential in future research and industry application.

**Keywords:** Transit real-time data; GTFS; waiting time; volunteered optimization.

1. **Introduction**

Mobile technologies are changing people’s life in different ways, also for transportation especially public transit. Real-time information (RTI) and real-time transit apps (RTA) are reshaping our way to take public transit (PT) system. With all the benefits of PT system, many people still felt reluctant to take public transit. Topmost reasons why people do not take PT are: long travel time especially long waiting time, lack of comfort, lack of control/certainty, and unreliability (Beirão & Cabral, 2007). Correspondingly, RTA can improve PT user experience in three aspects: strengthening control over the timetable, optimizing waiting time, and adapting to unreliability.

RTA provides users ability to comprehend the sophisticated timetable in a PT system, for both scheduled and real-time timetable. Paper timetable are inflexible and limited to schedule, and traditional phone call and text service are also limited to scheduled static timetable. RTA provide both scheduled and real-time support for PT users with smart phone through user-friendly interface.

RTA decreases the waiting time. Waiting time is a critical factor affecting people’s preference of transportation (Beirão & Cabral, 2007). RTA can plan best time for users to leave for the public transit based on the walking time and PT timetable. Ideally, RTA can diminish the waiting time to 0, which means as soon as users arrive at the stop, the bus arrives. Many studies investigated the waiting time reduction by real-time information: 91% percent of RTI users self-reported spending less waiting time in Seattle, 2010 (Ferris, Watkins, & Borning, 2010), and RTI users can save 2.4 minutes in Seattle, 2011 (Watkins, Ferris, Borning, Rutherford, & Layton, 2011) and 1.79 minutes in Tampa, 2014 (Brakewood, Barbeau, & Watkins, 2014) according to self-reported survey.

RTA adapts to PT system’s unreliability when optimizing waiting time (Brakewood et al., 2014). For all PT system, delay is inevitable. When delay happens, real-time transit apps (RTA) users can use their smart phone to get access to the real-time status of buses and plan their trips accordingly. With the help of RTA, users can have an anticipation about the on-time transit’s delay, thus postponing the departure time to reduce the waiting time at the stop. Many RTAs provide real-time trip planner to help users to schedule a best time to take public transit, such as Google Map and Transit.

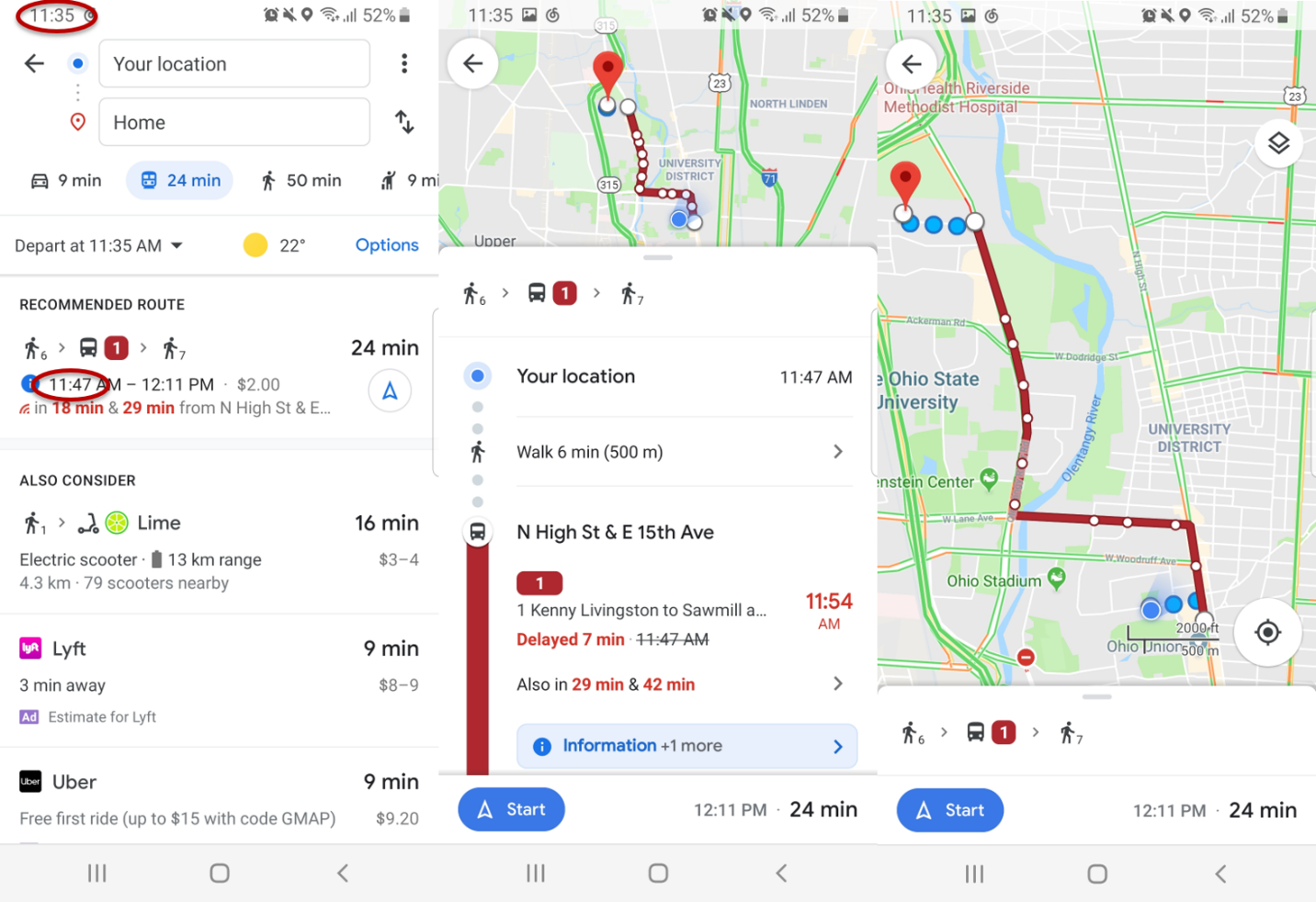


Figure 1 A typical interface of a real-time transit app.

However, besides the benefit RTA produces, the RTA trip planning strategies can lead to several undesirable results. The optimization of waiting time also comes with high risk of missing a bus. Figure 1 shows the changing trend of waiting time along with user’s arrival time at the stop, supposing the buses are all running absolutely punctually. The optimal point, where waiting time is 0, is exactly the point with highest risk of missing the scheduled bus, where the waiting time is the headway. Meanwhile, with the walking time increases, the instability of the PT trip’s real-time performance also increases. During the time when users are walking from home to the stop, a delayed bus can speed up to catch up the former delay, which can result in missing the bus for RTA users. Therefore, we would like to calculate and compare the missing rate for both non-RTA and RTA users.

Figure 2 Hypothetical waiting time's changing pattern along with user’s arrival time.

Nevertheless, even with more and more technology breakthroughs in this area, there are still economic and intellectual barriers for certain population to get access to the real-time public transit information. People who do not use smart phone application or do not own a smart phone cannot know the real-time status. These ordinary users have to plan their trips according to the schedule, their daily experience, or even random. Based on this fact, we would like to assess the average waiting time difference between non-RTA and RTA users and social justice issue behind the difference.

In conclusion, this paper is to assess and optimize the waiting time reduction and consequent missing risk between non-RTA and RTA users by consulting real-time bus feed. Based on the results, we can prove whether it is practically useful for users to follow different possible trip planning strategies derived by RTA’s trip planner, and how reliable different RTA strategies are. We will adopt several reasonable assumptions to build mathematical models and implement the models with General Transit Feed Specification (GTFS, real-time transit data standard) data provided by Central Ohio Transit Authority (COTA) in Columbus, Ohio.

1. **Literature review**

Analyzing the impacts of timely public transit information predates the development of contemporary RTI provided via webpages and smart phone apps. For example, Reed (1995) investigates signage and telephone’s real-time information’s impact on passengers’ waiting time (Reed, 1995). After the widespread application of smart personal devices, real-time information is becoming more accessible due to less expensive automated vehicle location system and the open data policy. Correspondingly, the body of literature is steadily growing and there are numerous studies investigating real-time information’s impact on public transit users (Brakewood et al., 2014; Brakewood, Macfarlane, & Watkins, 2015; Cats & Gkioulou, 2017; Ferris et al., 2010; Fries, Dunning, & Chowdhury, 2011; Liu, Shi, & Jian, 2017; Papangelis, Nelson, Sripada, & Beecroft, 2016; Watkins et al., 2011).

Brakewood and Watkins (2018) systematically review and categorize most research studying real-time transit information impact on the public transit system and users (Brakewood & Watkins, 2018), including waiting time, path choice, and ridership. They classifies studies across five dimensions: information media, mode of transit, location, methodology, and findings.

* 1. **Information media**

We can categorize research according to their information media, including static signage, telephone and text services, and smart phone application. Signage and at-stop displays can provide PT users useful information and reduce actual and perceived wait time (Dziekan & Vermeulen, 2006; Reed, 1995). Moreover, at-stop displays’ psychological effect is even more important: systems showing the next train or bus’s departure time can greatly release anxiety (Dziekan & Vermeulen, 2006). However, users can only get arrival information at the stop; this limits its effectiveness beyond reassuring the user since the decision to leave the origin has already occurred.

Telephone service and text services are more useful than signage since people can get access to the real-time information before deciding to travel to the stop or station (Reed, 1995). However, the provision of real-time transit information via the World Wide Web and smartphone apps has made this information more accessible and useful.

* 1. **Method**

Methods for analyzing the impacts of public transit RTI include survey-based methods, simulation models, and aggregate-level econometric analysis (Brakewood & Watkins, 2018). Survey-based methods include on-board surveys (Fan, Guthrie, & Levinson, 2016), before-after surveys (Chow, Block-Schachter, & Hickey, 2014), web-based surveys (Ferris et al., 2010), in-person surveys (Watkins et al., 2011), interviews and observations (Papangelis et al., 2016), and stated preference surveys (Liu et al., 2017).

**Survey-based methods**. Survey-based method is definitely the majority among all RTI impact studies. Surveys sample respondents with the same questions; they measure many variables, test hypotheses, and conclude temporal sequence from questions about past behavior, experiences, or characteristics (Neuman & Robson, 2014). Without the support of automatic real-time data, passenger surveys are the most direct methods to assess PT system use, especially for user experience and perceptions. Meanwhile, survey data can quantitatively assess different attributes using self-reported data (Neuman & Robson, 2014), which partially guarantees generalizability and authenticity. Besides, in contrast to the automatic generated data, the surveys’ data also point to users, instead of vehicles. The human-centered nature of survey data also guarantees its direct and close connection with human per se.

However, survey methods have their problems: Some survey-based methods rely on user’s self-reported information, which lacks precision and authenticity especially for non-cognitive value. Surveys measure the user’s perceptual estimation of the assessed value. Compared with value obtained by physical sensors, the self-reported information may be obfuscated and biased by threatening questions (Bradburn et al., 1979) and its self-evident nature (Goyder, 1986).

Survey’s sampling strategy and measurement is often questionable. First, the survey’s size, especially for some traditional data collection methods, is critically small due to high cost of data collection (Goyder, 1986). Second, some methods, especially IT (Information technology)-based methods, are often biased and dubious. Mail survey (Rossi, Wright, & Anderson, 2013), text/phone call survey, and internet-based survey (Wright, 2005) can significantly reduce the time and economic cost of the survey. However, these methods face a same problem: it is hard to access a representative sample (Rossi et al., 2013; Wright, 2005). For public transport system, not all users can get access to these services while the survey will only sample certain specific people.

**Simulation.** Mathematical simulation is often used to investigate and solve problems that are too difficult or costly to measure directly. For example, Cats and Gkioulou (2014) adopted an agent-based model to simulate the influence of PT reliability and real-time information on waiting time uncertainty (Cats & Gkioulou, 2017). Agent-based model simulation usually adopts several assumptions and represents the simultaneous actions and interactions of various agents. The simulation tries to imitate and predict the performance of a complex system such as PT system (Gkioulou, 2013). However, the effectiveness of the model is debatable, and the adopted assumptions could also be inconsistent with the reality.

Although the overall impact of RTI on waiting time is well explored, few studies investigated the variance (Brakewood & Watkins, 2018), including spatial and temporal patterns. Meanwhile, ironically, there are no studies using actual real-time transit data source due to the lack of these data and corresponding theory supports. We would like to address these gaps of waiting time problem in the real-time transit data context and implement it using actual real-time transit data.

1. **Methodology**

In this section, we first introduce the data source and corresponding manipulation processes. Moreover, we theorize the synchronization process during the procedure of transit and the concept of *delay reclamation*. Based on the synchronization theory, we propose and model several trip planning strategies (TPSs) during the process of decision making and optimize the RTA user’s strategy based on deterministic real-time data. We also calculate the waiting time difference between RTA users’ deterministic process and non-RTA users’ probabilistic process.

* 1. **Data**

Introduced by Google in 2006, the General Transit Feed Specification (GTFS) consists of two data stnadards: GTFS static and GTFS real-time expansion. GTFS static indicates the schedule data of a PT system in several separate tables (Google Developers, 2016). GTFS static is the current *de facto* standard for PT system schedules and PT geographic information (Google Developers, 2016). As a standard for open data, it is easy to share and access for the public, open-source programmer, and researchers. For this paper, we collected and organized all history schedule data in MongoDB and Python environment from Application Programming Interface (API) provided by Central Ohio Transit Authority (COTA) bus system in Columbus, Ohio.

Besides schedule data, GTFS real-time data provides vehicles’ geographic data with high temporal resolution. GTFS real-time includes two parts: buses’ location real-time data and the trip updates, which specifies the fluctuations in the real-time timetable (Google Developers, 2018). Unlike many traditional transit data, GTFS real-time data have large volume, high resolution, normalized standard, and full system coverage. For this study, we collected GTFS real-time data from API from COTA bus system with the frequency of 1 minute for more than 1 year.

* 1. **Synchronization**

Taking a bus could be conceptualized as a synchronization process between the walking trip to the target stop and the target bus’s *trip sequence array*. Trip sequence array is defined as the collection of trips running on the same route as the target bus in the direction of target bus.

Depending on user’s arrival time at the stop *t*, the actual bus that user will take can be different from the scheduled one. We use the same concept in the transfer synchronization process: *desynchronization degree* (DD), to measure the desynchronization between the bus and user at the stop. DD is an integer indicator that represent how many buses the user loses in the trip sequence array.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

If the user arrives at the stop between bus DD = n – 1 and bus DD = n, then the user will take bus DD = n.

The process of walking is linear: the users can strictly control the walking time by selecting their home departure time (HDT). For a user, the relationship between HDT and arrival time is linear.

Nevertheless, the actual real-time performance of the bus is non-linear: first, the users cannot directly control the boarding time by selecting their HDT, and the mathematical relationship between HDT and user’s boarding time is non-linear; second, the bus will not run at a fixed velocity and the expected time of arrival of bus at the stop is constantly changing. If the bus is delayed when the user departs home, during the walking time, the bus may catch up a part of delay by accelerating.

We therefore define the concept of *reclaimed delay (RD).* It is the time difference between the actual time of departure (ATD) and the expected time of departure (ETD) at the stop.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

Where: S is the collection of stops on the route, i is the target stop, t is the current time.

The reclaimed delay could be small but critical for RTA users: consequently, the RTA user will lose the bus and suffer waiting time penalty for a relatively long time. Thus, the synchronization of these two processes is highly unstable. Figure 3 shows a space-time diagram of the synchronization process.



Figure 3 Space-time diagram of the synchronization and desynchronization

* 1. **Non-RTI trip planning strategies**

A trip planning strategy (TPS) can be interpreted as a tactic for a user to plan and execute the transit trip. Assuming no disturbance on user’s walking and boarding process, different TPS has only one controllable factor to determine the actual waiting time: the time to leave home for the transit (home departure time, HDT). RTA relaxes the fixed timetable in a frequently delayed PT system, thus saving waiting time for RTA users. Depending on how to determine the home departure time, there are different trip planning strategies for both RTA and non-RTA users and their different purposes. Since we have access to the buses’ real-time information, if the home departure time is given explicitly, the calculation of TPSs’ real-time performance is deterministic.

In traditional incidence behavior studies, users are divided into two groups: timetable-dependent passengers, who are aware of the system schedule or empirical performance, and timetable-independent passengers, who are not aware of the system schedule and performance (Frumin & Zhao, 2012). These two user groups were introduced in the context of non-RTA experience.

**Arbitrary relaxation (AR)**. Before the time of smart phone, text, and public real-time information, under many circumstances, PT users were not necessarily planning their trips. A simple strategy is to walk to the stop and catch the subsequent bus arbitrarily. The major assumption is that the user’s arrival time is independent from the vehicle system (Frumin & Zhao, 2012).

Because the user’s decision-making process is random, it is reasonable to assume user’s HDT or user’s arrival time is evenly distributed among the headway between two buses. Traditionally, the average waiting time is the expectation of the random variable headway (Frumin & Zhao, 2012).

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Where: is the average waiting time, is the average headway, is the standard variance of headway.

However, since we have access to the deterministic real-time vehicle departure time, we can calculate the average waiting time as the mean of the departure time of target bus and its subsequent bus. Thus, the users’ home departure time is:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Where: is the median of the target bus and its subsequent bus. is the bus’s actual real-time departure, is a small constant time, is the very first prior bus’s actual real-time departure.

Theoretically, this strategy is not very efficient. We will also calculate the waiting time difference between AR and NR. This is another good benchmark: if a TPS’s performance is even worse than AR, we can say that it is not effective.

**Null relaxation (NR)**. Without knowing any information about the running status, users can still follow the schedule published to the public in advance. Or if a user has urgent affairs, such as work with strict timetable and medical emergency, she/he prefers earlier final arrival time than convenience. Under this circumstance, the user will follow the scheduled timetable of the PT system regardless of waiting time.

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Where: is the user’s home departure time (HDT), is the walking time from user’s home to the stop, is the scheduled bus departure time.

Consequently, the RTA user will not benefit from waiting time reduction. However, since no bus/train will leave earlier than the scheduled time, NR minimizes the missing risk. NR is another benchmark for waiting time reduction strategies, which has the lowest missing risk: if a TPS outperforms NR, it also means this TPS will outperformance most non-RTA users.

**Empirical relaxation (ER**). If a user can get access to the historical information, either from daily experience or a database, the user can relax the HDT based on the empirical average time/maximum time without any RTI.

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

where: is the user’s empirical arrival time for the bus trip

For the derivation of , different users may adopt different ER TPS. In general, there are two parameters to determine an empirical HDT: learning function and learning period. Learning function means the rules to generate the historical information, such as mean and maximum. Learning period is the available time period for the learning process. Ordinary users will not memorize all the historical information due to limited memory and limited access to the historical data.

For example:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

where: is the bus trip’s actual departure time on day *i*, and n is the learning period (days). This specific ER TPS’s learning function is average and its learning period is n days.

ER is a common non-RTA strategy. We also investigate waiting time’s relationship with ER TPS’s learning function (average/maximum) and learning period (1 days – 20 days).

* 1. **RTI-based trip planning strategies**

For the simulation of RTA trip planning process, most RTAs will directly use the ETDs provided by GTFS trip update for the buses’ real-time information (Google Developers, 2018; Transit app, 2019). The process of RTAs’ decision making consists of several steps: First, RTAs will provide different path choices for users and their HDT or users will find the desired bus trips/routes from the list in the RTA. Then, the trip update data will provide ETDs at the target stops for RTAs or users. Finally, RTAs or users will subtract estimate walking time and obtain estimated HDT. The estimated HDT is not constant; instead, RTAs will update it according to the real-time trip update data. Depending on the relationship between given estimated HDTs and the current time, the user will decide the actual HDT. Just like non-RTA TPSs, the essential part is the criteria to derive the actual HDT.

**Greedy relaxation (GR)**. Most transit users do not want to wait. Moreover, if a user encounters extreme weather events, especially cold weather and heavy precipitation, she/he may want to reduce the waiting time as much as possible. In practice, a greedy relaxation (GR) user will check the relationship between suggested HDT and current time by consulting RTA. She/he will only leave if RTA tells him so according to the real-time data, when the bus’s ETD at the stop is equal to or greater than walking time plus current time:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

Where: is the scheduled bus’s ETD at the stop given by RTA and real-time data, and is the current time when .

This strategy can achieve temporary optima. However, due to the instability of PT system, the missing risk of GR is also the highest. Due to the possible reclaimed delay, the bus will likely leave earlier than ETD. Even a smallest disturbance during the walking process or the bus running status during user’s walking can result in a missing bus using this strategy. Consequently, the user will suffer from a long waiting time penalty, which is almost equal to a headway, the largest possible waiting time.

**Prudent relaxation (PR)**. If a user would like to save waiting time and keep some degree of missing risk simultaneously, she/he may leave a short buffer of several minutes earlier than GR. Similar strategies were adopted in the context of scheduled time: some passengers will leave 2-3 minutes before scheduled service arrival in case of some unexpected “risky” events, such as delay by the elevators (Fonzone, Schmöcker, & Liu, 2015). These risk aversion reactions are usually defined as costs of unreliability (Frumin & Zhao, 2012).

Similarly, a prudent RTA passenger will leave a short buffer for the risk of missing bus and unexpected delay reclamation. This short time buffer, which is defined as *insurance buffer* (IB), trades some time to reduce missing risk, thus decreasing instability of the buses’ real-time performance. PR users’ HDT is:

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

where: IB is the insurance buffer.

Figure 3 also suggests the relationship between IB and reclaimed delay. RTA TPS is risky due to the existence of reclaimed delay, thus, the ultimate purpose of IB is to compensate for the reclaimed delay. Therefore, in the theoretical optimal PR scenario, an optimal insurance buffer should be exactly equal to the reclaimed delay to achieve 0 waiting time. Practically, IB can also relax the instability caused by the discontinuity nature of RTI feeds and some measurement errors.

Insurance buffer’s value is also a good indicator of the transit users’ risk attitude: It represents how much time the user is willing to gamble to gain the waiting time reduction. We define two extreme values of risk attitude: *risk-seeking* and *risk-averse*. *Risk-seeking* means the user would rather seek for the waiting time reduction regardless of the potential miss risk, which will possibly incur an additional large waiting time caused by desynchronization; *risk-averse* means the user would rather wait more time to avoid desynchronization. The less IB’s value is, the more *risk-seeking* and less *risk-averse* the user is.

In this sense, greedy relaxation (GR) can be viewed as the most risk-seeking TSP, since the user leaves no insurance to seek for the maximal expected waiting time reduction. Meanwhile, null relaxation (NR) can be viewed as the most risk-averse TSP, since the user is willing to follow the schedule to avoid desynchronization with longer expected waiting time. Prudent relaxation, otherwise, can be regarded as a balance point between the two polar, which can be called *risk-neutral* TSP. PR and GR can be categorized as a PR family, for GR is a special case of PR with IB = 0. With different IBs, each PR strategy can vary in actual waiting time. We would like to optimize IBs and find the best PR TSP with maximal waiting time reduction.

**Prudent relaxation with optimal insurance buffer (PR optimal)**. Previous research concentrated on PT system’s optimization in the stage of planning and operation. Due to the lack of authority and information access, ordinary users were rarely considered as a part of optimization process. However, with RTI, although users still cannot directly improve the real-time systematic performance (delay, ridership) of the system per se, optimization can be conducted in the individual level to reduce waiting time.

Correspondingly, we introduce the concept of volunteered optimization: despite PT systems’ instability and uncertainty, users with RTI can adapt and optimize each trip according to the delay and real-time information. Volunteered optimization is independent from the scheduled timetable; instead, it is based on the real-time status and decentralized. With the maximization of waiting time reduction in the individual level, the overall waiting time penalty will also be diminished. To achieve volunteered optimization, it is necessary to optimize each individual trip, which is from PR TPS family. Correspondingly, we need to find the PR TPS with optimal insurance buffer: *PR optimal*.



Figure 4 Space-time diagram of the PR optimal's trip

For PR or GR TPS, the users will plan their HDT according to the bus’s ETD instead of the real departure time. Here we define expected waiting time:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

where: is the ETD, t is the user’s arrival time at the stop, is PR strategy’s HDT, and is the walking time.

Combining (8) and (14), we have:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

For PR family, the expected waiting time is equal to the insurance buffer.

Figure 4 shows the theoretical relationship between users’ HDT and expected waiting time, missing risk, and waiting time.

Figure 5 Theoretical relationship between user HDT and average expected waiting time, average actual waiting time, and average risk.

To find the , empirically, we simulated the users’ real-time waiting time using different IB. We constitute the optimization problem in the following formula.

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

Subject to:

|  |  |  |
| --- | --- | --- |
|  |  | (12) |
|  |  | (13) |
|  |  | (14) |
|  |  |  |

Where: is the actual waiting time for the user live walking time from the stop who intends to catch trip . is the collection of all trips, and is the stops on trip , is the designated walking time range. is the actual departure time of target bus, and is the actual arrival time at the stop for the user. is the expected waiting time as well as the insurance buffer. is the expected departure time of the scheduled bus.

Figure 5 is the flow chart of PR optimization: We calculate the performance for all buffers for optimization, then we find the smallest waiting time and the corresponding buffer. If there are multiple smallest waiting time, designate the one with smaller buffer. After getting optimal buffers for every day and every trip and stop, we reduce all buffers into one day’s buffers by finding the maximum of the optimal buffers. In this way, we guarantee the optimality of obtained buffers: first, obtained buffers in each day have the least waiting time; second, obtained buffers are the smallest one among the buffers with the least waiting time; third, reduced buffers are the maximum buffer, which guarantees the synchronization for each day when recalculating the performance. In this way, we adopt a *risk-neutral* strategy: we are trying to find the smallest buffers while trying to keep synchronized for most trips.

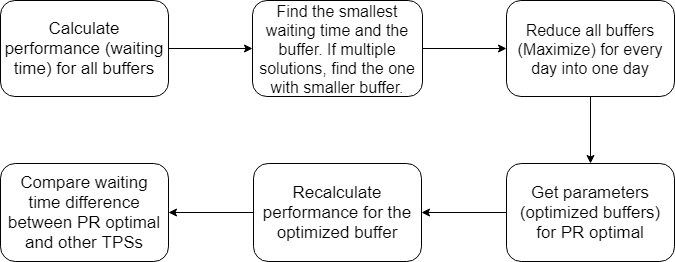


Figure 6 Flow chart of PR optimization algorithm

For the first dominating step of the optimization process, the computation complexity is , where b is the number of possible buffers, d is the number of dates, t is the number of GTFS trips, s is the number of stops in a trip, w is the number of walking time, n is the number of the GTFS feeds for the stop. Thus, the total complexity is polynomial and of high power. The optimization process of produces massive number of parameters . We minimize waiting time over , which will have a different IB for each day, each trip, each stop, and each walking time. In practical, to reduce computation load, we calculated the AWT for every stop on COTA bus route No. 2 for different IB, from 0 to 300 seconds with interval of 10 seconds. We also parallelized the outmost loop to improve computation performance on a workstation with 40 virtual CPU cores.



* 1. **Measures**

We would like to measure the difference of waiting time and risk of missing a bus/train between the RTA users and non-RTA users. Therefore, we define two measures: the missing risk (MR) and average waiting time (AWT).

**Missed bus risk.** The missed bus risk measures the probability of missing a bus based on the TPS relative to the actual performance of the transit system:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

Where: n is total number of trips;

Here we define missing a bus/train as a phenomenon that the actual bus’s desynchronization degree is larger than 0. This also means that the user takes a different bus after the scheduled bus.

**Average waiting time**. Average waiting time measures the expected wait time across all trips based on the TPS. We start by defining the actual wait time:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

Where: is the actual waiting time, is the bus’s actual real-time departure time, and is user’s arrival time at the stop. is user’s HDT, and is user’s walking time from home to the stop.

So, the bus’s actual real-time departure time is:

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

Thus:

|  |  |  |
| --- | --- | --- |
|  |  | (12) |

Equation 12 proves that the only factor that user can control and can affect waiting time is .

The mathematical expectation of waiting time’s distribution across all trips is the average waiting time:

|  |  |  |
| --- | --- | --- |
|  |  | (13) |

where: n is total number of trips.















**Waiting time difference**. For non-RTA users who has no access to the real-time data, the most rational and practical strategy is NR or AR. And for RTA users, the most beneficial strategy is PR with an optimal insurance buffer. We introduce the difference between NR's waiting time and ’s waiting time:

|  |  |  |
| --- | --- | --- |
|  |  | (20) |

In practice, we will calculate for GR,, ER, AR, and compare these strategies’ efficiency.

* 1. **Implementation**

We collected GTFS real-time trip update data and corresponding GTFS static schedule data of the COTA (Central Ohio Transit Authority) bus system in Columbus, Ohio from January 2018. The data is stored in a MongoDB database. The GTFS trip update contains the ETD and ATD for every stop and every minute. We developed the algorithm in Python environment and maintained our smart transit database. We constructed a generic GTFS real-time database with 300 GiB volume in MongoDB and also generated other auxiliary databases in terabyte level in total. We simulated the working process of RTA and user’s trip planning process by calculating the walking distance between the user and the target stop.

With the support of MongoDB databases, we moreover developed a web-map visualization interface using JavaScript and Leaflet. The web-based geographic information system can be generalized and expanded into any PT systems with GTFS real-time support.

1. **Analysis**

In this section, we focus on the geographic and temporal analysis of buffer, waiting time, and waiting time difference between different TPSs. For the study area and target, we choose COTA bus No. 2 from Southeast to Northwest from Feb 2018 to Feb 2019. The bus route has two schedules: the frequent one originates from blue circled stop (frequent originating stop) in Figure 7 with headway of 15 minutes or better, while the standard on originate from red circled stop (standard originating stop) in Figure 7 with headway of 15 – 30 minutes (COTA, 2019).

* 1. PR optimal

**PR optimal’s parameters.** We first calculated and analyzed the optimal insurance buffer (IB) for prudent relaxation optimal strategy. Figure 7 (left) shows the geographic distribution of PR optimal’s insurance buffer. It shows an extreme geographically diverse pattern for IBs. So, we cannot assign a single IB for all stops and different users with different walking distance.

Figure 7 (right) demonstrates the miss risk for the PR optimal strategy, and it shows an unnatural high miss risk cluster near the originating stop. This is because the lack of real-time data in the very beginning of bus trips. The real-time information will not be available until the bus leaves the originating stop. Also, sometimes the data will not be updated promptly because the data’s update frequency is 30 seconds. For certain trips, by the time the real-time information is updated, the user already loses the bus. Under this circumstance, buffer will not help improve the miss risk of such trips since IB’s effectiveness depends on accessible real-time information. Meanwhile, users who live far from the stop will have higher miss risk and will suffer from more waiting time.

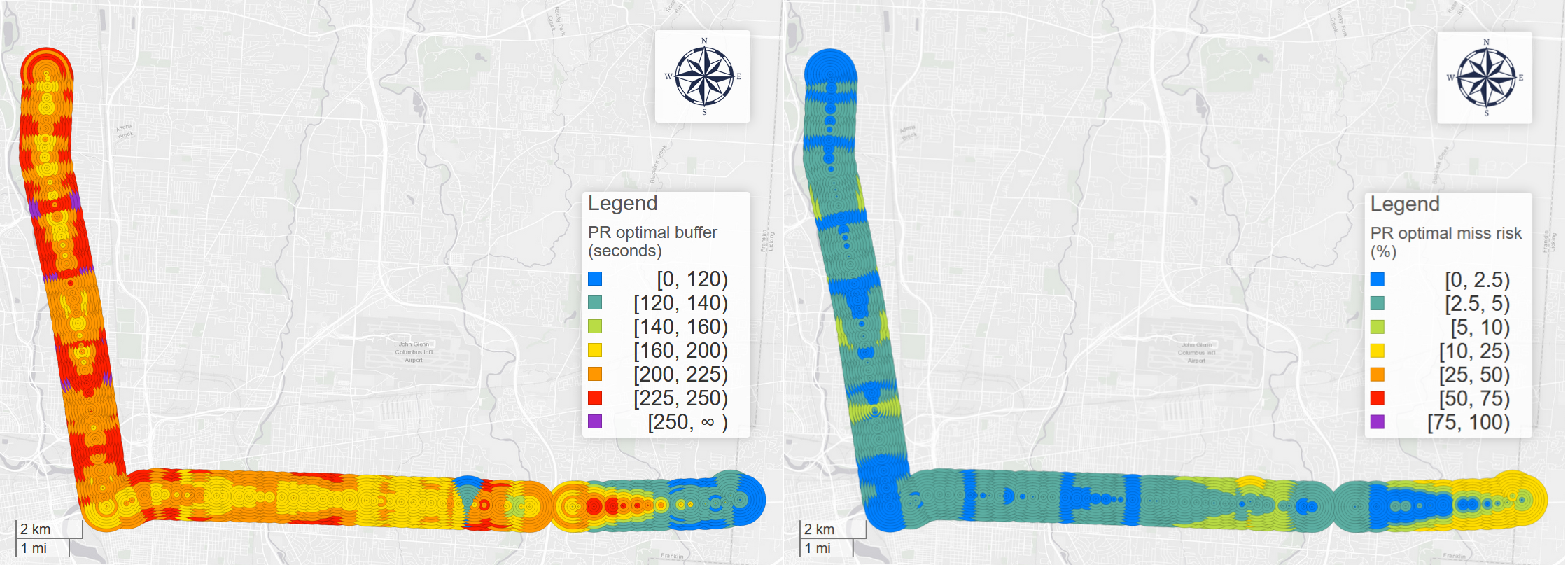


Figure 7 PR optimal’s insurance buffer for each stop and walking time in COTA bus route No. 2 from Southeast to Northwest in 2018.

**PR optimal’s waiting time.** Figure 8 (left) shows PR optimal’s average waiting time on COTA bus route No.2 from Southeast to Northwest. Noticeably, there are two significant high clusters near the two originating stops (red circled and blue circled) in the standard and frequent No.2 bus schedule.

Out of the two high clusters, the standard originating stop (red circled) has higher waiting time. First, it is because of the high miss risk, which can also be observed in Figure 7 (right) due to the lack of real-time information in this area. Besides, the headway near the standard originating stop (blue circled) are larger in the bus schedule. These two factors contribute to increasing the waiting time, therefore, the standard originating stop (red circled) have larger waiting time than the frequent originating stop (blue circled).

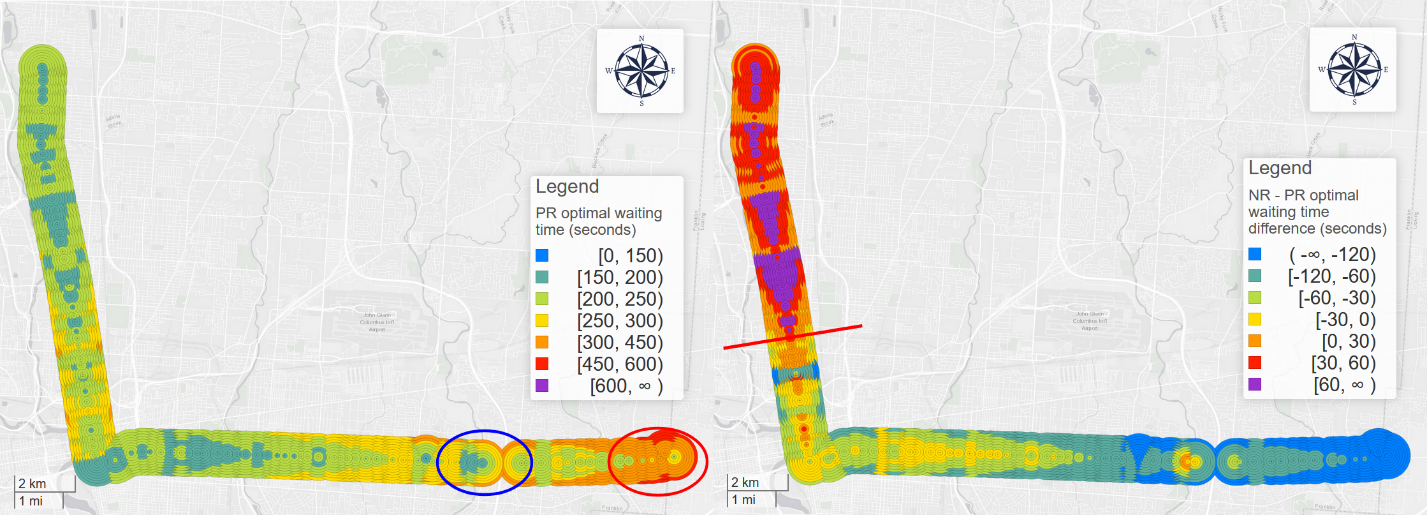


Figure 8 PR optimal’s average waiting time (left) and NR – PR optimal waiting time difference (right) for each stop and walking time in COTA bus route No. 2 from Southeast to Northwest in 2018.

* 1. Other TPSs

Figure 9 visualizes the waiting time of GR (top left), NR (top right), AR (bottom left), ER (bottom right) and Figure 10 shows the waiting time difference between AR/ER/GR and PR optimal. First, arbitrary relaxation has an average waiting time of half of the headway. Therefore, we can observe a drastic change between standard schedule and frequent schedule due to the increase of headway. Intuitively, the performance of AR should have been the worst among the TPSs.

However, arbitrary relaxation has the best waiting time of 494 seconds among the three TPSs. In fact, greedy relaxation, as the only RTA TPS, has the second worst waiting time of 594 seconds and empirical relaxation using average has the worst average waiting time of 636 seconds. This proves that RTA users without proper advice could wait significantly longer than even arbitrary relaxation.

The reason is that GR and ER (using average) will be more likely to desynchronize when catching bus, thus have large miss risk. Statistically speaking, AR’s probabilistic distribution’s mean is half of the headway before the schedule; however, the mean of ER (using average)’s probabilistic distribution is bus’s departure time’s mean for each day. Therefore, when an ER arrives according to the average time, a considerable portion of distribution will be desynchronized with the user arrival time. Within bus departure time’s distribution, those desynchronized portion which falls before the average departure time, which is also the ER user’s arrival time, will incur a time penalty equal to almost a full headway.

Likewise, GR’s mean is also near or even after the departure time; also, due to its deterministic nature, GR could result in systematic higher chance of desynchronization, making its performance nearly equal to ER’s. This can also be observed from GR and ER’s miss risk.

Figure 10 (bottom right) shows the miss risk difference between GR and PR. GR’s average miss risk of is 59.56% and ER’s average miss risk is 56.29%. In contrast, PR optimal’s average miss risk is 3.39%. This also proves the high risk of desynchronization for GR and ER (using average). Therefore, in the sense of both average waiting time and miss risk, GR and ER (using average) are not effective TPSs.

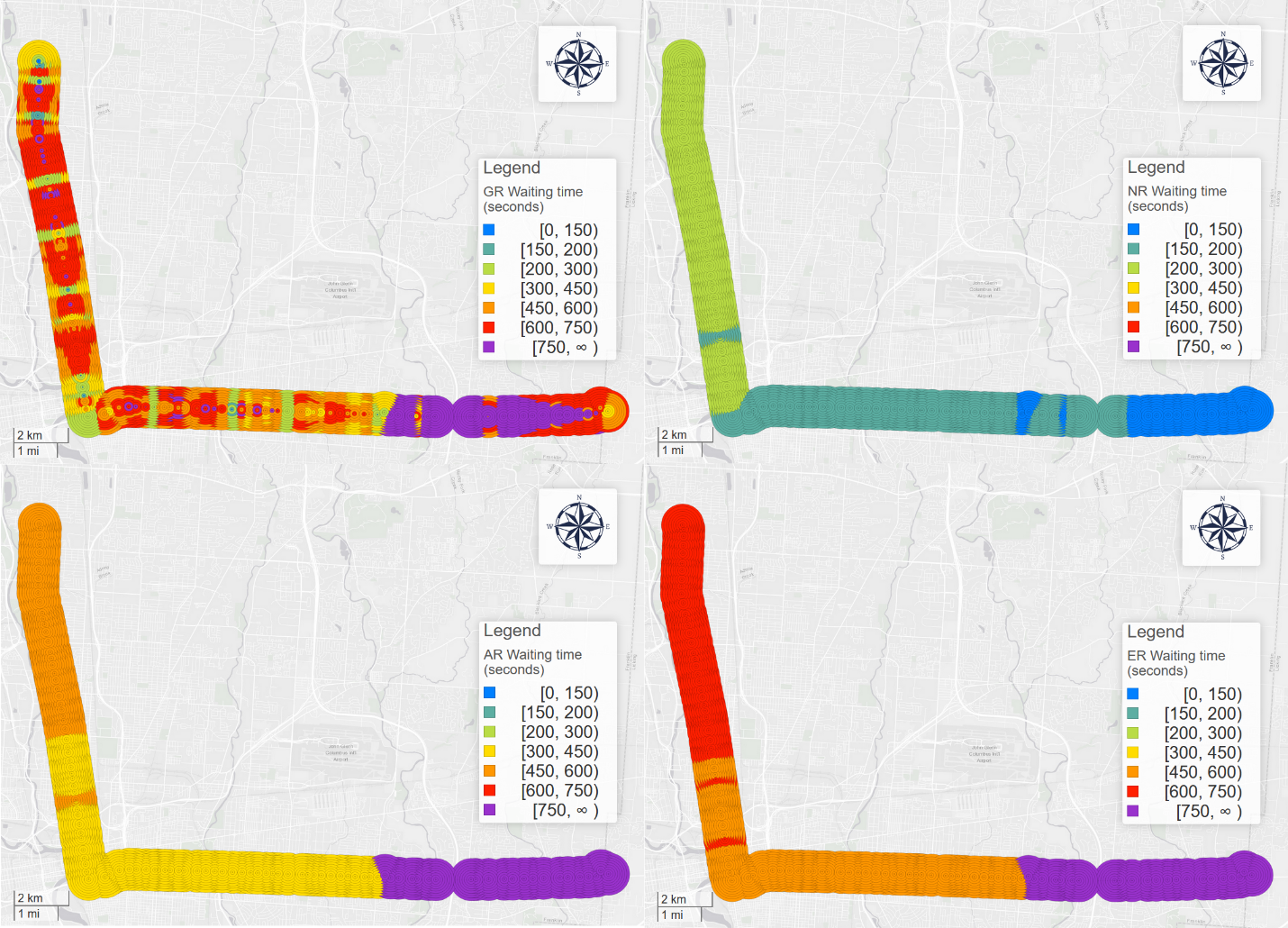


Figure 9 GR (top left), NR (top right), AR (bottom left), ER (bottom right)'s waiting time pattern

**Difference between PR optimal and NR.** Figure 8 (right) shows the average waiting time difference on COTA bus route No. 2 from Southeast to Northwest. The differences represent the distinction between performance of best RTA users (PR optimal) and best non-RTA users (NR), respectively. We can observe that PR optimal does not outperform NR for all stops. In fact, for most stops, especially for those stops in the upstream near the originating stops, NR’s performance is much better than PR optimal.

For PR optimal which is independent from the schedule, we can observe a relatively less oscillating pattern for all stops, expect the originating stops with exceptional high waiting time. For most stops, waiting time keeps a relatively low level, but we can hardly observe a stop with an even lower waiting time. This is because IB reduces desynchronization risk when the user catches the bus, but it also makes people wait more time for these synchronized trips. In conclusion, too large (risk-averse) or too small (risk-seeking) buffers all impair the effectiveness of PR optimal, but too small buffers will especially result in desynchronization and suffer more waiting time. Moreover, for stops near the originating stops, RTA users will be more likely miss the bus due to the lack of real-time data, thus waiting even longer time than the schedule.

However, for NR, its waiting time is exactly the delay of each bus: the average delay starts from 0 while it accumulates and propagates along the route with fluctuations, as shown in Figure 9 (top right). Consequently, due to PR optimal’s relatively large waiting time in the upstream stops, NR users will wait less than PR optimal users. On the other hand, for areas with significant delays, PR optimal will outperform NR.

The comparison moreover proves that PR optimal cannot achieve absolute optimality, instead, the upstream and downstream stops show polarized patterns. We divide the stops into two groups at stop “North High Street & Euclid Avenue” shown as a red line in Figure 8 (right). In average, PR optimal users had to wait 227 seconds compared to the NR’s 206 seconds. However, for upstream stops, people who observe PR optimal had to wait 58 seconds more than the people who follow the schedule; while for downstream stops, PR optimal users saved 47 seconds compared with the NR users.

As a result, for best performance, RTAs can combine PR optimal and NR into one new TPS: designate HDT with the maximum of PR optimal’s and NR’s HDT. In this way, if PR optimal’s HDT is even less than NR’s, users will be told to leave home according to the schedule to avoid wasting time caused by too large buffer.

* 1. Geographic patterns

Although many figures above visualize different variables, they share a similar geographic pattern. We investigate the geographic patterns from three directions.

[Vertical – Delay propagation]

Park, Mount, Liu, Xiao, & Miller (2019) investigated COTA bus system’s on-time performance and public bus delay propagation using GTFS real-time data (Park, Mount, Liu, Xiao, & Miller, 2019). The delay propagates via bus system’s network. The paper also presents the pattern of propagating delays and models it with an exponential model:

|  |  |  |
| --- | --- | --- |
|  |  | (21) |

Where: is the difference of delays at stop and its subsequent stop (stop ’s propagating delay). is the original delay at stop p. is the transit route-based distance from p to q. is the delay constant (Park et al., 2019). The propagating delay is exactly the reclaimed delay. As a result, in the vertical direction, formula 21 also applies to the reclaimed delay and the ideal insurance buffer. The miss risk, waiting time, and waiting time difference are all correlated with the delay propagation/reclamation. However, formula 21’s may be not strictly applied to route No. 2 and the presence of delay reclamation is not significant. We visualized each variable’s actual vertical changing pattern in the Analysis section.

[Diagonal – Contour lines]

An obvious diagonal pattern that can be observed in the map is parallel diagonal contour lines. In the map, for each point on the contour line, their values are the same. Most maps above prove the presence of the contour lines. We can conceptualize the contour lines in a smaller scale: two neighboring circles at two subsequent stops with the same value. It is the smallest unit of the contour line. These two circles’ walking time difference is nearly equal to the bus’s running time between the two stops. With more neighboring circles with the same value, the contour lines are formed.

There are two reasons for this phenomenon. First, for the two neighboring circles, essentially these users in the two areas are aiming for the same bus but at different bus stops; second and most importantly, users in the two areas will get a same real-time data simultaneously and plan their trips according to the very same real-time data feed, thus arrive at the target stop at the same time. Therefore, if no disturbance or acceleration between the two stops, the two trips are the same only with different walking distance. This also suggests that the gaps on the maps are the point where delay oscillates, and the bus is running in the constant speed in the areas with smooth contour lines. The contour lines’ formation can also be understood in another temporal sense: after the bus receives a disturbance at a stop, the disturbance will persist and spread outwards when the bus is moving to following stops.

[Horizontal – Walking time influence]

Figure 12 illustrates the relationship between waiting time and walking time and Figure 13 shows the relationship between miss risk and walking time. For PR optimal, the longer walking time is, the longer will the user wait and the riskier the user will be to miss the bus. This supports the claim that the longer distance the user lives from the stop, the more unstable their trip becomes. During the longer walking time to the stop, the bus could be more likely to accelerate to catch up the delay, making RTA users miss the target bus.

For GR, due to its high miss risk, longer walking distance larger than 120 seconds will in fact improve the performance. The fact that miss risk is initially at a high level makes longer walking time a similar role as insurance buffer: longer walking time will destabilize the synchronization process and make more GR’s desynchronized trips synchronized again.

Figure 12 Each TPS's waiting time's relationship with the walking time

Figure 13 Each TPS's miss risk's relationship with the walking time

* 1. Temporal patterns

In this section, we conducted several temporal analyses based on different resolutions.

* + 1. Daily patterns

Figure 14 visualizes ER, AR, and NR’s daily average waiting time for each day from February 2018 to February 2019. ER and AR’s waiting time patterns show strong periodicity: Saturdays have significantly larger waiting time because of sparser bus schedules, and the average waiting time oscillates with the cycle of a week. For AR, the average waiting time is about 450 seconds, which is the half of 15 minutes (900 seconds); for ER, the average waiting time is the largest among the three non-RTA TPSs. The NR’s average waiting time, which is also the average delay, is more irregular but still correlated with ER and AR’s waiting time.

Figure 14 ER, AR, and NR's daily average waiting time.

Figure 15 visualizes NR, PR optimal (the average of 10 walking time scenarios), and GR (the average of 10 walking time scenarios)’s daily average waiting time. It also strengthens the claim that PR optimal cannot achieve absolute optimality: for most days, PR optimal didn’t outperform NR for global average waiting time.

Similarly, PR optimal and GR also shows strong periodicity: Saturdays are significantly larger than other days of week. Besides, we observe that the waiting time becomes larger after September 3rd, 2018. It could be because of major route or real-time adjustment made by the transit authority.

Figure 15 NR, PR optimal, and GR's average waiting time.

Figure 16 Each TPS's waiting time.

* + 1. Day of week

Figure 17 visualizes each TPS’s waiting time on each day of week. For ER, AR, GR, and PR optimal, they share similar changing patterns: Saturday is the highest day, and Friday is the lowest day; however, for NR, Saturday is the lowest and Friday is the largest day. This phenomenon moreover demonstrates the negative correlation between delay and RTA TPSs’ relative effectiveness: the more the system is delayed, the more effective RTA is.

Figure 17 Each TPS's waiting time on each day of week.

* + 1. Hour

We also analyzed the hourly patterns for each TPS. Figure 18 visualized the hourly average waiting time for ER, AR, GR, PR optimal, and NR. For high headway hours like 4:00 to 7:00 and 21:00 to 24:00, ER, AR, GR, and PR optimal all have higher waiting time, since the price of missing a bus will dramatically increase. Although the overall performance (NR’s waiting time/ average delay) of the system during these hours is better than normal hours, on the contrary, users using these TPSs will not benefit but will suffer from buses’ large headway. This also suggests that delay is not the only and absolute standard to assess a system’s performance, instead, the user’s experience is more important. This is especially true in the context of real-time transit apps.

Like the global average waiting time, for most hours, the sequence is ER>GR>AR>PR optimal>NR. However, there are several exceptions: In general, ER and GR share an extremely similar changing patter, however, for high headway hours, GR’s waiting time is larger than ER. This suggests that GR is more sensitive to the headway variation.

For PR optimal specifically, we can observe two valleys from the curve: 7:00 – 9:00 and 15:00 -19:00, which are exactly the morning and afternoon rush hours. Again, this also suggests that PR optimal works better with higher delay in the system.

Figure 18 ER, AR, GR, NR, and PR optimal's hourly average waiting time.

* 1. Marginalized stops (I will consider deleting this part since I have talked about the lack of real-time information in section 4.1.)

We can observe that the first few start stops near the originating stop, which can be called *marginalized stops*, have a significant larger waiting time difference. It suggests that RTA strategies may be less effective at marginalized stops. In these areas, the walking time could be larger than the running time from the start stop to the target stop. These stops are marginalized, in the geographic, temporal, and social justice sense.

Theoretically, these users’ HDT could be earlier than the bus’s start time at the originating station, which make the process even more risky. When the users should be heading for the stop, the bus has not left the originating stop, where users cannot have access to the real-time data in advance. Figure 19 shows marginalized stops’ waiting time difference for different routes in COTA bus system. This phenomenon universally exists in the system. The best option for users in this area is adopting NR and following the schedule.

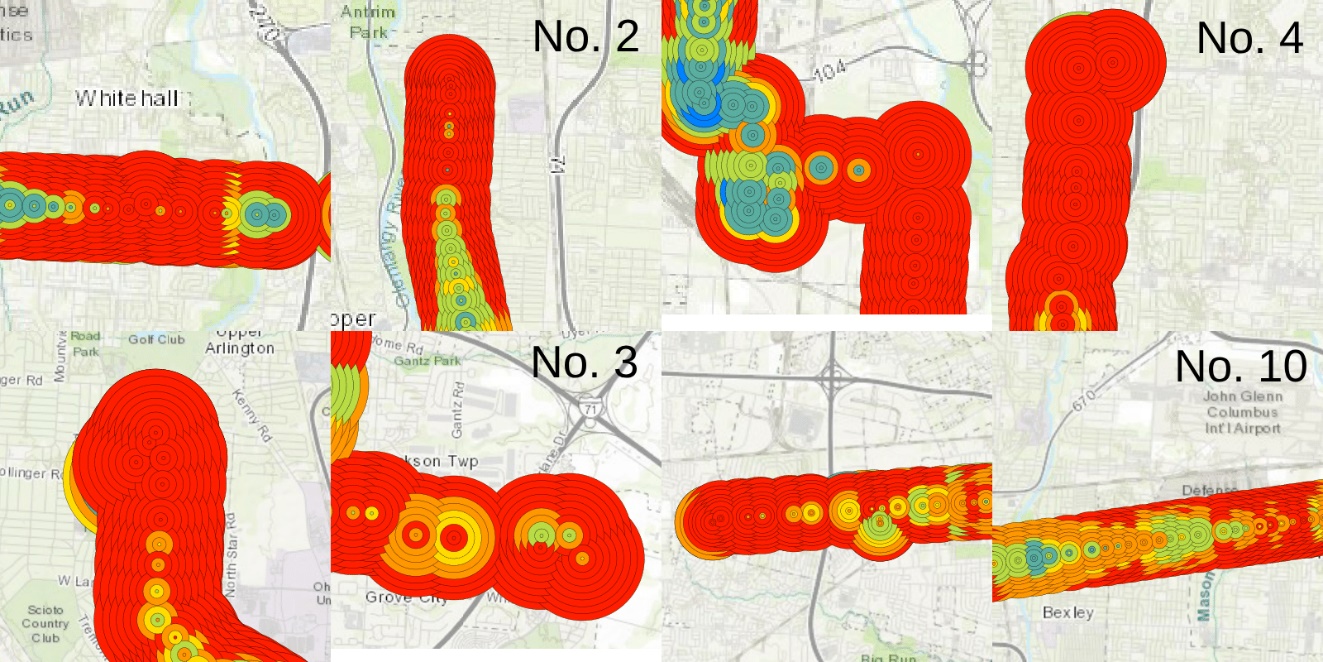


Figure 19 Marginalized stops' waiting time for COTA route No. 2, 3, 4, and 10.

1. Conclusion

Real-time information (RTI) and real-time transit apps (RTA) can significantly decrease transit users’ waiting time (Brakewood & Watkins, 2018). However, few studies systematically investigated waiting time reduction’s spatiotemporal patterns and optimized the system performance from users’ perspective. In this study, using GTFS real-time data, we first developed *volunteered optimization* theory and calculated RTA and non-RTA users’ real-time performance. We theorized different trip planning strategies (TPSs) under different scenarios and user groups. We also introduced the proportion of different user groups and risk attitudes for each TPS. Then, we optimized RTA users’ *prudent relaxation* TPS with optimal insurance buffer, so that RTA users can minimize waiting time in the individual level. Based on the PR optimal strategies, we calculated the waiting time difference between different TPSs and conducted geographic and temporal analysis in different directions and resolutions. Moreover, we observed the presence of *marginalized stops*.

The empirical results and the volunteered optimization system can provide vital information for transit users, planners, and real-time transit apps. With more access to real-time data, transit system planning should not only engage with the schedule but also real-time performance; RTA development should not only engage with real-time performance but also empirical performance; passengers’ trip planning should not only engage with empirical performance but also all the information above. To achieve these three goals, future RTAs should combine schedule, real-time, and empirical information into one, with corresponding computation and networking support. For example, add pre-calculated insurance buffers to GTFS data so that RTA trip planning results inflect PT system’s empirical performance.

Reference:

Beirão, G., & Cabral, J. A. S. (2007). Understanding attitudes towards public transport and private car: A qualitative study. *Transport Policy*, *14*(6), 478–489.

Bowman, L. A., & Turnquist, M. A. (1981). Service frequency, schedule reliability and passenger wait times at transit stops. *Transportation Research Part A: General*, *15*(6), 465–471.

Bradburn, N. M., Sudman, S., Blair, E., Locander, W., Miles, C., Singer, E., & Stocking, C. (1979). *Improving interview method and questionnaire design: Response effects to threatening questions in survey research*. Jossey-Bass San Francisco.

Brakewood, C., Barbeau, S., & Watkins, K. (2014). An experiment evaluating the impacts of real-time transit information on bus riders in Tampa, Florida. *Transportation Research Part A: Policy and Practice*, *69*, 409–422.

Brakewood, C., Macfarlane, G. S., & Watkins, K. (2015). The impact of real-time information on bus ridership in New York City. *Transportation Research Part C: Emerging Technologies*, *53*, 59–75.

Brakewood, C., & Watkins, K. (2018). A literature review of the passenger benefits of real-time transit information. *Transport Reviews*, 1–30.

Cats, O., & Gkioulou, Z. (2017). Modeling the impacts of public transport reliability and travel information on passengers’ waiting-time uncertainty. *EURO Journal on Transportation and Logistics*, *6*(3), 247–270.

Chow, W., Block-Schachter, D., & Hickey, S. (2014). Impacts of real-time passenger information signs in rail stations at the Massachusetts Bay Transportation Authority. *Transportation Research Record*, *2419*(1), 1–10.

COTA. (2019). 2 N HIGH / E MAIN. Retrieved June 3, 2019, from https://www.cota.com/wp-content/themes/gotravel-child/PDF/lines/2.pdf

Dziekan, K., & Vermeulen, A. (2006). Psychological effects of and design preferences for real-time information displays. *Journal of Public Transportation*, *9*(1), 1.

Fan, Y., Guthrie, A., & Levinson, D. (2016). Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security. *Transportation Research Part A: Policy and Practice*, *88*, 251–264.

Ferris, B., Watkins, K., & Borning, A. (2010). OneBusAway: results from providing real-time arrival information for public transit. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1807–1816). ACM.

Fonzone, A., Schmöcker, J.-D., & Liu, R. (2015). A model of bus bunching under reliability-based passenger arrival patterns. *Transportation Research Procedia*, *7*, 276–299.

Fries, R. N., Dunning, A. E., & Chowdhury, M. A. (2011). University traveler value of potential real-time transit information. *Journal of Public Transportation*, *14*(2), 2.

Frumin, M., & Zhao, J. (2012). Analyzing passenger incidence behavior in heterogeneous transit services using smartcard data and schedule-based assignment. *Transportation Research Record*, *2274*(1), 52–60.

Gkioulou, Z. (2013). Evaluating the impact of waiting time uncertainty on passengers´ decisions.

Google Developers. (2016). GTFS Static Overview | Static Transit | Google Developers. Retrieved March 8, 2018, from https://developers.google.com/transit/gtfs/

Google Developers. (2018). Trip Updates. Retrieved April 8, 2019, from https://developers.google.com/transit/gtfs-realtime/guides/trip-updates

Goyder, J. (1986). Surveys on surveys: Limitations and potentialities. *Public Opinion Quarterly*, *50*(1), 27–41.

Jolliffe, J. K., & Hutchinson, T. P. (1975). A behavioural explanation of the association between bus and passenger arrivals at a bus stop. *Transportation Science*, *9*(3), 248–282.

Liu, Y., Shi, J., & Jian, M. (2017). Understanding visitors’ responses to intelligent transportation system in a tourist city with a mixed ranked logit model. *Journal of Advanced Transportation*, *2017*.

Luethi, M., Weidmann, U., & Nash, A. (2007). Passenger arrival rates at public transport stations. In *86th Transportation Research Board Annual Meeting (TRB 2007)*. Institute for Transport Planning and Systems, ETH Zurich.

Neuman, W. L., & Robson, K. (2014). *Basics of social research*. Pearson Canada Toronto.

Papangelis, K., Nelson, J. D., Sripada, S., & Beecroft, M. (2016). The effects of mobile real-time information on rural passengers. *Transportation Planning and Technology*, *39*(1), 97–114.

Park, Y., Mount, J., Liu, L., Xiao, N., & Miller, H. J. (2019). Assessing public transit performance using real-time data: spatiotemporal patterns of bus operation delays in Columbus, Ohio, USA. *International Journal of Geographical Information Science*, 1–26.

Reed, T. B. (1995). Reduction in the burden of waiting for public transit due to real-time schedule information: a conjoint analysis study. In *Pacific Rim TransTech Conference. 1995 Vehicle Navigation and Information Systems Conference Proceedings. 6th International VNIS. A Ride into the Future* (pp. 83–89). IEEE.

Rossi, P. H., Wright, J. D., & Anderson, A. B. (2013). *Handbook of survey research*. Academic Press.

Transit app. (2019). How we shrank our trip planner till it didn’t need data. Introducing public transit’s fastest, tiniest, offline-capable trip planner. Retrieved December 5, 2019, from https://medium.com/transit-app/how-we-shrank-our-trip-planner-till-it-didnt-need-data-84984ca56663

Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S., & Layton, D. (2011). Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A: Policy and Practice*, *45*(8), 839–848.

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, *10*(3), JCMC1034.