Comments from the editors and reviewers:

1. Reviewer 1

This paper addresses a relevant topic and offers some insightful results. It does need however in my view to undergo substantial revisions before becoming suitable for publication. In the following, I note 9 comments which I consider major and thereafter list minor comments.

Major comments:
1.1. The paper is currently unnecessarily lengthy. I think the essence of the paper and its substance are such that do not require 16 figures and 29 pages. The paper requires some considerable trimming.

Will summarize later

1.2. The authors overstate their results. In essence, the obviously naive approach of planning to arrive the same second as the bus is expected to arrive can be expected to often (and even a majority of cases if there is even a slight systematic bias towards over-predicting the remaining time for arrival) result with missing the bus. I find the so-called the insurance buffer to be the more innovative element in this study. Note its affinity with the notion of 'hidden waiting time' which refers to a high (e.g. 95th) percentile of the expected distribution.

This is a fair comment. We cut down the size of the analysis part and only keep the essence of the results. We answer the questions raised by this comment in following aspects:

- Is it trivial to study the naive greedy tactic?

We also made some clarification in our responses to the question

Greedy tactic is the default tactic that most transit planning apps suppose their users to use. The default scenario of using a transit planning apps is: the app will provide a home departure time / leaving time calculated from real-time data; many apps will not consider the risk of missing a bus during talking and use a greedy tactic to suggest leaving time. This means that many apps are expecting users to arrive the same second as the bus is expected to arrive. Meanwhile, if users do not realize the risk and choose to believe the results of the transit apps, this user will automatically use a greedy tactic. We think this is one of the contributions of the paper: transit apps which use greedy tactic to calculate their recommended home departure time are not trustworthy, and we provided a systematic way to prove this claim based on the actual real-time data the transit apps are using. We also added more justification in section 3.4.4.

-

1.3. Section 2.1 is redundant. I suggest the authors instead review the methods of quantifying the impacts of RTI on waiting times and thereafter report their findings. Also the relevance of the paragraph on surveys and their costs in 2.2 can be removed without any loss of information, unless the authors intend to specifically comment on those conducted in the context of RTI and waiting times.

This is a good suggestion and we removed the former section 2.1 and organize the literature review as the reviewer suggested.

1.4. The literature review I miss analytical approaches such as the studies performed by a study published in this journal ("Evaluating the added-value of online bus arrival prediction schemes") and a related study published in the Journal of ITS ("Real-time bus arrival information system: An empirical evaluation") by the same authors. These studies also considered the impact of RTI on waiting times along the line, as well as function of the prediction horizon which is equivalent to the analysis in relation to walking time performed in this paper. Moreover, the comparison with a waiting time resulting from following the schedule has also been performed. The authors should better acknowledge how their work relates to previous work.

Thank you for pointing this out. We added these useful references to the paper and we believe the literature review is now more complete.

1.5. The presentation of the method can be dramatically shortened. GTFS and APC are by now very standard data sources in transit research. The notion of "reclaimed delay" is also not new. The impact of initial delays on running times further downstream has been extensively studied with conflicting evidence, see for example:
El-Geneidy, A. M., J. G. Strathman, T. J. Kimpel, and D. T. Crout. 2006. “Effects of bus stop consolidation on passenger activity and transit operations.” Transportation Research Record
El-Geneidy, A. M., J. Horning, and K. Krizek. 2011. “Analyzing transit service reliability using detailed data from automatic vehicular locator systems.” Journal of Advanced Transportation.
Cats, O. 2018. "Determinants of bus riding time deviations: Relations between driving patterns and transit performance". Journal of Transportation Engineering.
Similarly, departure strategies that have been reported elsewhere like AT can be described briefly.

Thank you very much for pointing this out and we found these literature very useful. The notion of “reclaimed delay” and arbitrary tactic are definitely not new. The main contribution of the

1.6. The authors refer to measurement error and schedule recovery efforts as an explanation for an over-estimation of bus arrival time. Even in the absence of those, an over-estimation could simply result from traffic conditions including short signals and skipping stops (no boarding and alighting passengers).

This is a very useful complement and we changed the definition of reclaimed delay in section 3. We also considered these factors in our analyses and calculation. The over-estimation scenario can be a major source but not the only source for desynchronization. We can categorize and decompose all sources into two classes: the first one is the measurement error, which does not come from the vehicle's performance change. The second one includes all the vehicles' performance change, including all the mentioned incidents and delays above. These errors can be detected by GTFS records during our calculation by calculating the arrival/departure time at each stop. In other words, we actually could not detect the exact reason from the buses' performance why the desynchronization happened; for example, we cannot distinguish whether it is because of the bus driver's accelerating or a short signal. But this won't affect the actual results shown in this paper.

The comment is very useful and important, so we extended the concept of "reclaimed delay": we made it not exclusive for driver accelerating and add the instances the reviewer mentioned.

1.7. A critical point is that IB is introduced only in PT but none of the other TPSs includes an element that is conscious of risk-taking. ST can also include an IB term, i.e. avoiding just missing the bus. This applies also to ET.

This is a good question. We will provide some analyses in this comment.
Conduct buffer analysis to ST and ET.

However, we did not

1.8. Notations throughout the manuscript are sloppy. For example, introducing if definitions without indicating the value taken otherwise. There are also other matters, needs to be revisited carefully.

We reworked all of our notations and equations with simpler and more intuitive expressions.

1.9. Some of the conclusions may not be transferable, can the authors please reflect on that? For example, the performance of ST obviously depends on the on-time performance of the service under consideration (in particular, the share of early arrivals). The spatial pattern discussed in 4.3.3. is also clearly caused by the deterioration of the on-time performance further downstream.

Minor comments:

1.10. Suggest to shorten the title. After the question mark can simply have only "An empirical analysis"

1.11. TPS is mentioned in the last paragraph of Section 1 but has not been introduced yet

1.12. Add axes titles in figures 1 and 2

1.13. Broken references in section 3.3

1.14. Frumin and Zhao (2012) is not the original source for Eq. 3. Please refer to the original contributor.

1.15. The description accompanying Figure 3 is not sufficiently clear, please revisit.

1.16. Please revisit also the last paragraph in the conclusions, it is not clear to me what is meant by this.

2. Reviewer 2

This paper presents an analysis of a single bus route in Columbus, Ohio to explore the impacts of different passenger trip planning strategies, including those using real-time information, on passenger wait times. Overall, I found the manuscript to contain numerous noteworthy flaws. Specifically, the analysis relies on some unusual assumptions that may be driving the results; furthermore, the scope of the analysis is limited to a single bus route in a single city, limiting the generalizability of the findings. Moreover, the authors have not validated their theory and findings with real world behavioral data, such as from surveys or focus groups. In light of these weaknesses, I recommend significant revisions to the paper. My specific comments are detailed below corresponding to the page number (since there were no line numbers in the manuscript).

2.1. Page 2, Paragraph 3

How did the authors come up with the idea that RTI apps can diminish waiting times to zero? Have the authors conducted a survey/focus groups/interviews of riders to demonstrate that this "greedy" strategy is something riders actually do? Numerous prior studies of waiting times have been cited in the literature review section of this paper, and they all included reality high average wait times (e.g., Watkins et al. found 9.23 minutes for RTI users compared to 11.21 minutes for non-users). It seems unrealistic to expect riders to minimize their wait times to zero.

This is a very good question and one of the very reasons we choose to investigate the RTI's impact. We also mentioned the same issue in our response to the comment 1.2. We will respond the comment by following several sub-questions.

- *How did the authors come up with the idea that RTI apps can diminish waiting times to zero? Numerous prior studies of waiting times have been cited in the literature review section of this paper, and they all included reality high average wait times (e.g., Watkins et al. found*

9.23 minutes for RTI users compared to 11.21 minutes for non-users). It seems unrealistic to expect riders to minimize their wait times to zero.

It is all true and we made corresponding changes in the text to avoid confusion. As the reviewer mentioned in this comment and according to our results in the paper, RTI apps can never *always* diminish actual waiting times to zero and we did not suggest that they can. The "zero wait time" refers to its expected waiting time, instead of actual wait time in Watkins et al. "Zero expected waiting time" means: if everything does not change during the user's walking and the arrival time keeps the same as the ETA when the user was scheduling the trip, the greedy tactic user's waiting time will be zero. As the reviewer pointed out, this is clearly a very ideal scenario.

However, this greedy tactic is already adopted by many transit apps to calculate their suggested leaving time before this paper. We summarized the implied principles of greedy tactic and recreated the algorithm in this paper, but essentially the main purpose is to simulate how transit planning apps generally plan the trips for users, instead of creating a new algorithm or a new rule. It is very common for many apps (Google Map, Transit app) and open source trip planning projects (OpenTripPlanner) to assume and expect the user arrives at the exact second/minute when the bus arrives. There is no "waiting time" or "buffer time" in a trip suggestion in the interface of these apps; they will always show a suggested leave time, which is the expected arrival time subtract the walking time. Therefore, there trip planning apps are essentially following the greedy tactic. Some of the reasons are:

- The naïve and simplistic nature of greedy tactic: it requires no optimization like empirical tactic or prudent tactic, which will reduce the calculation burden for the calculation in the fly. GTFS data is also naturally design for the calculation: all the ETAs are already pre-calculated by the GTFS provider.
- No optimization also means no warranty. Therefore, when people miss the bus, people will not blame the optimization algorithm.
- The temporary optimal solution: as we explained in the former "zero expected waiting time" part, in the moment of trip scheduling, greedy tactic can ideally achieve the 0 waiting time. If assuming the bus performance being stationary, this is the temporary optimal solution.

However, as we shown in the paper, because the future is unpredictable (reclaimed delay, including bus accelerating and short signals) and the updated data's high interval (discontinuity delay), greedy tactic has a high risk of missing a bus, and missing a bus is highly penalized by longer waiting time. The empirical analyses also show that in fact the greedy tactic's actual performance is very terrible.

- Have the authors conducted a survey/focus groups/interviews of riders to demonstrate that this "greedy" strategy is something riders actually do?

As we discussed in the last section of the response, greedy tactic is adopted by many transit apps to calculate their suggested leaving time. Many apps (for example, Google Map and Transit) assume and expect the user arrives at the exact second/minute when the bus arrives. If the user follows the suggestion of these apps, even if she/he does not realize the risk of missing the bus, the user is following the greedy tactic.

In this paper, what we did is to summarize the implied principles of greedy tactic and recreate the algorithm adopted by the apps. The main purpose is to simulate how transit planning apps generally plan the trips for users, instead of creating a new rule. In fact, this is one of the contributions of our paper as we define the greedy tactic in a strict mathematical manner, which is not explicitly discussed before.

As for the issue of whether people will actually follow the suggested greedy tactic:

- First, it is certain that some transit app users will follow the suggestion, because this is the default and primary trip planning strategy provided by the apps. This is especially true for some new app users who are not aware of the high risk of greedy tactic.
- Second, greedy tactic can be a very good benchmark to measure an upper bound of the performance, regardless of people's actual usage rate. Greedy tactic is a special case of prudent tactic when the insurance buffer equals 0. As the most risk-seeking prudent tactic, it is important to investigate its performance.

Therefore, since many transit apps tend to use greedy tactic to calculate the suggested time, it is necessary and meaningful to investigate the greedy tactic's actual waiting time. We acknowledge the value of behavioral validation, however, as we indicated in the end of the paper, the behavioral validation by survey is a limit of this paper and a future direction of the research due to its length and exceeding contents.

For more clarification, please refer to our response to the comment 2.later.

2.2. Page 3, Section 2.1

The authors do not differentiate between perceived versus actual wait time differences in the literature review, which is an important distinction in prior research on the impacts of real-time information. Please add a brief discussion.

Will do

2.3. Page 6, Paragraph 1

The authors claim that APC data is more accurate in terms of arrival/departure time at each stop compared to GTFS-realtime. I found this surprising. How did the authors come to this finding? What analysis did you conduct to demonstrate the accuracy? Additionally, did you verify a sample of the data with real world observations (e.g., ride the bus and manually record the stop times, then compare them to APC and GTFS)? Please justify.

This is another great question and we made several adjustments in the paper to avoid confusions. Here, we are not discussing about the precision or accuracy in the spatial sense, such as the accuracy of GPS points. Here, we are referring to their *temporal accuracy*. Similar to Firmani et al. (2016)'s definition, we define it as: how accurate is the measure's recorded time compared to the actual time of event occurrence. It represents the systematic error caused by the temporal delay of measurement.

GTFS real-time data is a regularly updated data, which means there is always an interval between the measured time and the actual event time. In our analyses, the frequency of updating is

However, APC data is produced in an on-demand manner, which means the measured time is always the same as the actual event time.

2.4. Page 6, Paragraph 2

COTA's data is updated once per minute, which seems quite long compared to many other agencies. For example, the MBTA in Boston updates their bus location data every 5 seconds (see https://medium.com/@sjbarbeau/introducing-the-gtfs-realtime-validator-e1aae3185439). This is likely an important data limitation from COTA that is driving some of your results (discussed more later). Please add discussion of typical update/refresh times from other transit agencies.

Page 17, Figure 7 (we merged the two comments due to their relevance)

- Figure 7 shows high sensitivity to the 60 second update of real-time data, which, as previously noted, seems to be a reality high value (e.g., the MBTA updates every 5 seconds). Is there a way you can test the sensitivity of this in your model? At a minimum, it should be discussed as a drawback of the case study of COTA, as it appears to be driving the results shown in Figure 7 and may not apply to other agencies with better real-time data.

This is a very good point. We address this issue in following aspects:

- An overview of some transit systems' GTFS real-time trip-update data updating frequency

Here we show the update frequency of all publicly available transit systems in the US that provide GTFS real-time feed from OpenMobilityData.org (OpenMobilityData 2020). We used the GTFS real-time validator (Center for Urban Transportation Research @ USF 2020) to measure the update frequency of each GTFS real-time feed.

Transit system	Update interval (secs)	Transit system	Update interval (secs)
MBTA	~5	Go Metro, Cincinnati	~30
Community transit	~10	DCTA, Denton, Texas	~30
CATA, Lansing, MI	10 – 20	VIA, San Antonio	~30
MST, Monterey, CA	10 – 20	HART, Tampa, FL	~30
RTC, Southern Nevada	10 – 20	LTD, Eugene, OR	~30
Votran, Daytona Beach, FL	10 – 20	Metro Transit, Madison, WI	~30
ART, Arlington, VA	20 – 30	MTA Maryland	~30
Big Blue Bus	20 – 30	RTA, riverside, CA	~30
Calgary Transit	~30	Capital metro	~60
BART	~30	CT Transit, Hartford	>60

Among 20 transit systems we could successfully test the update frequency, 12 of them have update interval larger than 30 seconds as of May 2020; in fact, MBTA is the only known transit authority that releases such high-frequency GTFS real-time trip-update data feeds in the United States. Moreover, in Google's official GTFS playbook webpage, the GTFS real-time data is expected to have update frequency around 1 minute (Google Developers n.d.). These statistics shows that the majority of the transit systems in the United States still face non-trivial continuity delay in 2020 large than 30 seconds; the discussion of impact of long update interval larger than 30 seconds is still important for most transit systems.

Moreover, these statistics were calculated in May 2020. Each transit system will gradually increase the update frequency with better equipment and more experience. For example, we know from COTA that they has upgraded the update frequency for the recent two years. Therefore, it is very likely that many transit systems had a larger update interval back in 2018.

- The actual data update frequency of the transit planning apps

What users see on their smartphones are not the trip-update data, but results generated by the transit planning apps. The update frequency of the GTFS real-time data is different from transit planning apps update frequency.

Several examples can be given by using the Transit app, which is a popular transit planning app with millions of users. We used a regular Android phone and an iOS tablet to randomly select several routes and different time in a day for testing purposes. We observed the information update frequency on the interface is lower than the data per se. For example, MBTA, the transit systems with the highest update frequency, can still have update intervals from 15 seconds to 1 minute for most routes shown on the actual smartphone app interface, despite theoretically the data are served every 5 seconds. We also witnessed similar phenomena in

Capital Area Transportation Authority (CATA) in Lansing MI and Monterey–Salinas Transit (MST) in Monterey, California. This suggests that the actual information update can be lower than the ideal data update frequency.

- Further implication of higher update interval

Despite the large interval is still common for most transit authorities, the question raised by the comment is still important, because we will witness more transit systems with GTFS trip-update data of higher update frequency in the near future. However, as for COTA system, we could not get data of higher resolution for the same time period, therefore the potential impact of higher frequency still remain largely unknown. Therefore, it could be a limitation for this paper and could be a very good topic for future studies.

2.5. Page 6, Last Paragraph

The authors make the assumption that the walking process "is linear with respect to distance." They later explain that the only way to change the walking time is to depart the home at a different time. However, the assumption of constant walking speed seems highly unrealistic. If a rider sees a bus approaching and they think they might miss it, they are very likely to speed up and potentially even run to meet the bus. Indeed, a study by Dziekan and Kottenhoff (2007) of the subway in Stockholm, Sweden observed passengers entering subway stations and counted the number of passengers running and walking when RTI signage outside the subway station was on, and the results reveal that significantly more people run when the RTI signage was on rather than when the signage was off. In light of this real world evidence, the assumption of constant walking speed seems very unrealistic, and the authors should test the sensitivity of their results to altering this assumption.

Thanks for pointing this out. There are several reasons that we chose to use linear walking process and insurance buffer ahead as the primary assumption:

- Running is not a viable and desirable option for everyone.

Running after seeing the bus approaching can be possible for some younger people, but it is not viable other passengers such as senior people, disabled people, people with luggage, and parents with children. And the obstruction of buildings and trees, the condition of pavements, and the weather all make running a less likely option for most people. Even if a passenger can run to the stop, compared with planning an insurance buffer ahead, passengers may find it exhausting, stressful, and thus undesirable.

- Running should not be and is not the default assumption for transit planning apps.

Because running is not viable option for many people, the transit planning apps' planning logic should not naturally assume everyone should do that. In fact, most transit planning apps will not suggest a "running" phase or suggest people when to run during the walking phase when planning the trip. Therefore, since we are simulating the results provided by the actual transit planning apps and the behavior of people using these apps, we will not primarily assume people and apps will do this.

However, we know some people will accelerate after they see the bus approaching and some buses will wait for the passengers when the driver see the passenger. Therefore, per the reviewer's suggestion, we added a sensitivity test to test the results. The saved time cannot be too long for two reasons: first, human cannot run significantly faster than the usual walking speed; and second, the average sprint distance cannot be very long, since the obstruction of building and trees and the route number on the bus usually won't allow the passenger see a bus in a very long distance. We use the average speed of 3 m/s as the running speed (Chertoff 2018). We consult the letter height visibility table; for a route number on a regular US bus of 20 cm (7.8 inches), the ideal readable distance is about 21 (70 feet) – 24 (80 feet) meters. We will use 24 meters as the sprint distance. In our study, we use 1.4 m/s the standard walking speed for all the calculation. Therefore, the maximum saved time is 24/(3-1.4) = 15 seconds. We conducted the sensitivity test based on the extra 15 seconds for relaxation and add the results to the paper.

2.6. Similar to the previous comment, another recent study by Ferris et al. (2010) found that RTI provided on mobile devices may impact a passenger's decision of where to board the transit vehicle, which would impact the passenger's walking distance to access transit. On a survey of RTI users conducted in Seattle, Washington, 78% of respondents reported they were more likely to walk to a different stop based on RTI (Ferris et al., 2010). Can the authors test this (e.g., changing where to board) in their modelling framework?

Thanks for pointing this out. In our original analysis, there is a

2.7. Page 9, Table 1

How did the authors arrive at these 5 trip planning strategies? Similar to my previous comment, have the authors conducted a survey/focus groups/interviews of riders to demonstrate that these are strategies riders actually use?

Thanks for pointing this out. We will revisit each trip planning strategy (TPS) one by one to argue that each of them is actually used by people without conducting a dedicated inductive survey for justification of using these trip planning strategies. We will prove our point by deductive reasoning.

- Non-real-time trip planning strategy (Schedule, arbitrary, and empirical tactic)

Schedule tactic are the default trip planning strategy that how people will use public transit. Transit system has schedule and people are expected to follow it. It is most certain that people will use this. Arbitrarily leaving home for buses is also another very common TPS, if not the most common. It is most certain that people will use arbitrary tactic. People will memory the historical arrival time and find the earliest one or the average to decide when they will leave for the bus. It is also most certain that people will use the empirical tactic. Non-real-time trip planning strategies' usage is well proven by the common sense and daily experience.

- Greedy tactic

We justify the investigation of greedy tactic in this paper in several aspects:

First, as our responses to the comment 2.1 point out, greedy tactic is adopted by different trip planning apps and algorithms and we are not the first one to come up with it. It is common for the suggested trip plans generated by the apps to follow greedy tactic. Therefore, we can confirm that some people will follow the suggestions thus use the greedy tactic by deductive reasoning. We believe these facts and deduction can be a good evidence to answer the question raised by the reviewer about whether people will use greedy tactic: some people will surely follow the suggestion, although it is largely unknown about the exact ratio of these people.

Second, the necessity of investigating the performance of greedy tactic does not lay on whether or how many people use it. The fact that many apps, especially the most popular ones, adopted this strategy is our primary motivation to measure its performance. Just like what we asked in the title: does real-time information (and real-time information apps) reduce waiting time? If apps and current popular trip planning algorithms are systematically suggesting a trip plan with very poor performance based on real-time information, it is extremely necessary and important to measure its performance, including how bad it is and how bad it can be compared to other traditionally used non-real-time trip planning strategies. Moreover, it would be valuable to find a good way to improve the algorithm. All of these will be based on the investigation of the greedy tactic.

- Prudent tactic

Similar to greedy tactic, there are some facts supporting the actual usage of prudent tactic. We will justify our endeavor to investigate prudent tactic in following aspects:

First, users can actively use prudent tactic. As we mentioned in the section 3.4.5, it is a common strategy to avoid risk of missing a bus by leaving earlier than the suggested time. This is especially true for two kinds of people: the experienced users and users who are less familiar with the transit system, such as tourists and new residents. Experienced users will notice the unreliability of the suggested time thus leave earlier. New users will feel more insecure thus leave earlier than the expected time to avoid risk.

Second, users can passively use prudent tactic. We mentioned in the paper that greedy tactic is a special case for prudent tactic with insurance buffer =0. We already demonstrate most apps' default trip planning strategy is greedy tactic; therefore, any trip plan that the user chooses to use is a special form of prudent tactic. For example, if the suggested leaving time by the app using greedy tactic is 5:00pm and the user chooses to leave at 4:58pm, even if the person has no specific intention to plan an exact insurance buffer, the plan that user uses is in fact prudent tactic with buffer = 2 minutes. In this sense, people will certainly use prudent tactic, as long as they consult the real-time data and have an offset before the suggested time. This process can be perceived as a very primitive and coarse prudent tactic; and it is also why we should systematically discuss a more advanced version with well-optimized insurance buffers for specific time and space.

Last, just like greedy static, the necessity of investigating the performance of greedy tactic does not lay on whether or how many people use it. As we discussed in first two reason, for transit systems and app providers, it is extremely important to consider the performance retrospectively and provide an alternative and improved solution to the current greedy tactic. In this paper, we introduce prudent tactic and the prudent tactic optimal. It is certain that the prudent tactic optimal won't be used by anyone right now, but it would be better than the greedy tactic. And if current apps can adopt our solution or develop better solution to solve the insurance buffer and use them as the algorithm to generate suggested route, more people will benefit from this.

In conclusion, from the facts, common sense, and published results that we know, all tactics are indeed used by at least some people. For certain variety of prudent tactic and the prudent tactic optimal, we are sure that no one are using it right now and this is the very reason we are proposing a new algorithm: people will benefit from using it as it is adopted by more apps thus more passengers. On the other hand, it can be a very good potential for future studies to expand the topic to the inductive perspective, because the quantitative composition of users' trip planning strategies is still largely unknown, even if we can infer that people are using them. We thank the reviewer again for providing a useful insight.

2.8. Page 10, Equation 3

- It should be noted that the average waiting time formula for random arrivals is generally only applied to high frequency transit routes (e.g., headways less than 10-15 minutes).

This is very true for formula 6 and we added relevant clarification. However, we only give formula 6 to introduce the traditional calculation method. We did not use headway and its standard variance in our calculation. This is one of motivation to use the high-resolution data without major stochastic assumptions. The actual formula we were using is formula 7: we will find the bus before the targeted bus and produce the average of the two buses' time.

2.9. Page 3, Equation 5

- The authors state that the "bus will rarely if ever leave a stop earlier than the scheduled time." Did the authors verify this statement empirically, such as comparing the GTFS schedule to GTFS-realtime? Drivers occasionally do run "hot." Please justify the assumption that they don't.

Thanks for pointing this out. We did not assume that buses never run "not" in our calculation. The best answer to this question must be that we present a risk of missing bus for the schedule tactic. The missing risk of schedule tactic represents the chance for buses to leave earlier than the schedule, which is exactly the scenario of "running hot". From table 2 in the main text, we can see that the mean of schedule tactic's risk of missing bus is 6.28%. If we assume no buses will run "hot", the number will be 0.

As for the authenticity of this claim, the risk of missing bus is the lowest among other strategies; and 6% is a rather low absolute value. Moreover, in COTA's official on-time performance policy, buses are not permitted to run ahead of schedule for normal buses (COTA 2019). Even if a bus leave 1 second earlier than the schedule, the bus is not considered "on-time".

2.10. Page 11, Figure 3

- What data was used to create the visualization shown in Figure 3?

We implemented the empirical tactic of different memory and calculate the performance based on the same GTFS-APC data we collected. We did not use an external data source or reference. We added relevant clarification in the main text.

2.11. Page 13, Equation 9

- The authors state that "for PT family, insurance buffer should be at least equal to the expected waiting time." Please explain why this is the case. It wasn't clear to me.

2.12. Page 14, Last 2 Paragraphs

- The authors state that there is a large computational burden to conduct the analysis, so they only selected one bus route for the analysis. This greatly limits the generalizability of the research. Instead of using 1 year of data, why not use 1 week of data and run the analysis for multiple bus routes? I strongly encourage the authors to consider a larger geographic sample.

Thanks for mentioning this. In the following answers, we are going to answer the following several major questions raised by the comment.

- Will the selected bus route 2 limit the generalizability of the research?

In this response, we select 6 representative major routes: COTA route 1, 2, 5, 7, 8, and 10. All of these routes have different directions and wide spatial and temporal coverage in the city of Columbus. The six routes account for

Just as the reviewer suggested, we select the time period from a typical week from 7/15/2018 – 7/21/2018, when there was no major event like football games and extreme weather.

Figure 1 and Figure 2 shows the six routes' prudent tactic optimal's waiting time.

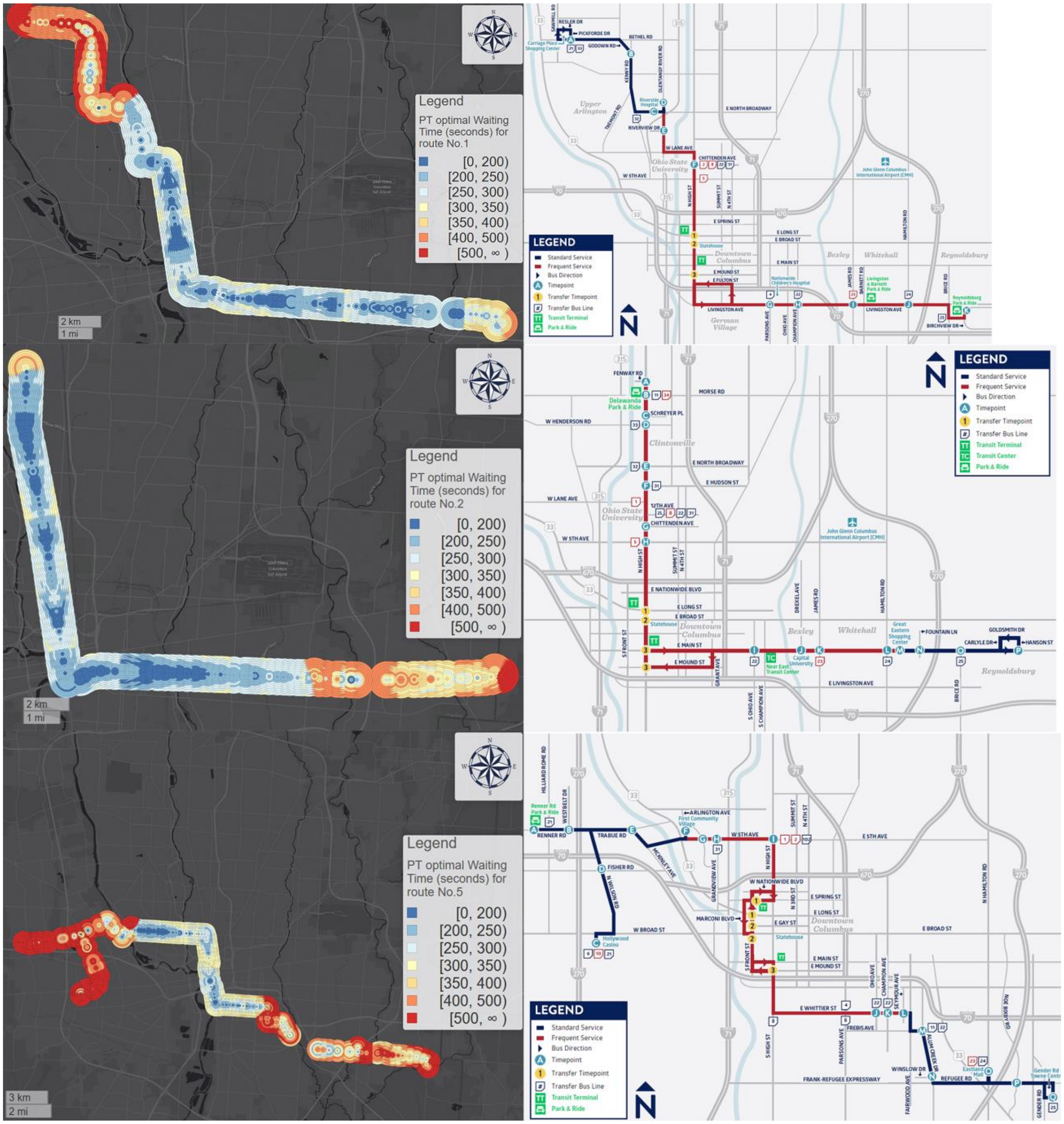

Figure 1: Route 1, 2, and 5's PT optimal's waiting time and route map.

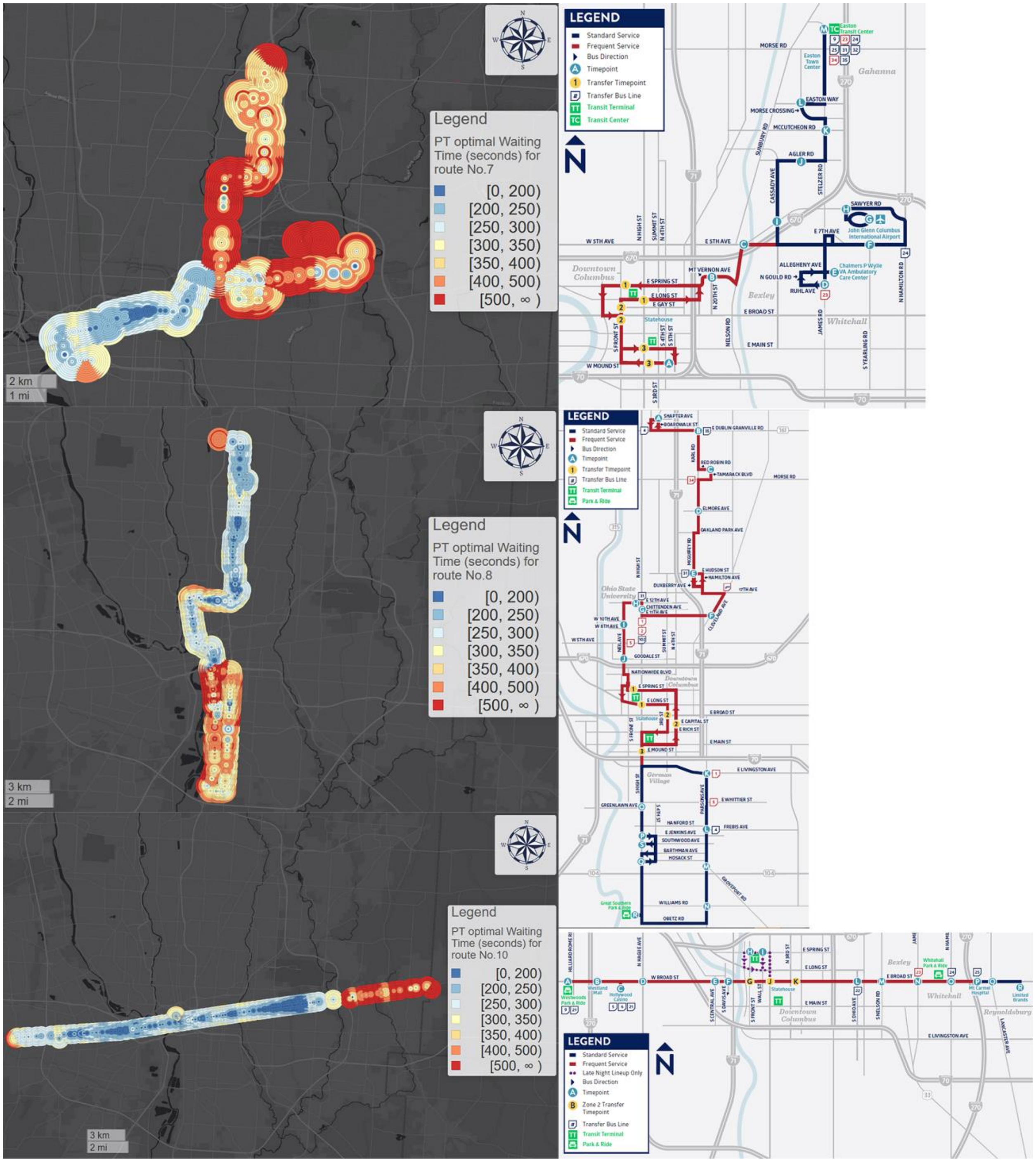

Figure 2: Route 7, 8, and 10's PT optimal's waiting time and route map.

2.13. Page 21, Figures 11 and 12

- I found the GT results counterintuitive, and I would have expected them to be similar to the PT results. I suspect this is likely due to the assumptions built into your models (e.g., constant walking speed so riders won't run for an approaching bus and the long 60-sec threshold of real time updates).

Yes, it is counterintuitive, however, we show that this counterintuitive phenomenon is justifiable. We explain that 60% of US transit system still have non-trivial update interval larger than 30 seconds, and MTBA is actually the only transit system that provides such as high frequency in US in our responses to the comment 2.4. We also explain why running is not possible for many people and should not be a default assumption for trip planning apps in our responses to comment 2.5. The sensitivity test also say a lot of things. This is the main contribution of the paper: we prove that the current algorithm has a very bad performance.

On the other hand, the poor performance of greedy tactic does not mean that real-time information is 100% useless. The majority of the problems are due to the naïve trip planning strategy and, just like the reviewer pointed out, the low update frequency. This is the second main contribution of this paper: we summarize the definition of prudent tactic and insurance buffer and introduce a systematic and sustainable method to optimize the insurance buffer. By using a better and adaptive trip planning strategy which combined the advantages of real-time data and empirical analyses, we can make real-time information more useful for passengers in future.

2.14. Page 22, Figure 13

- For the route shown in Figure 13, how many timepoints are there along the route? Where are the timepoints located? How do the timepoints relate to your findings (in this figure and the subsequent maps of the route)? Please add a discussion of timepoints.

Will do

2.15.

- For Figure 13, please make sure the colors in the legends correspond to the same numerical values. It is difficult to compare GT with ST, AT and ET since the ranges for the colors are different. This comment also applies to subsequent figures.

Will do

Minor Comments

2.16. On page 9, please fix the reference errors to the figures/tables.

Will do

2.17. The authors introduce many new acronyms throughout the paper, which can be confusing for readers. Please try to limit the use of acronyms that are not commonly found in the prior literature (e.g., consider removing IB, TPS, HDT, etc.).

Will do