Response to the Reviewer's Comments

Dear Reviewers and Editors,

Thank you very much for your helpful feedback in the previous round. In this revision, we have tried our best to address the concerns from the reviewers. For your convenience, we are submitting this review response along with our revised manuscript to explain our revisions. In the following, we will first itemize major changes made to the manuscript, and then we will also provide a point-to-point review response to both meta-review and each reviewer.

In this revision, we have primarily made the following changes. For each change, we have assigned a symbol which related to the corresponding review comment.

1. (M3) Perform a brief evalutaion on our speed estimation algorithm with a smaller dataset which contains both cell phone data and GPS data. We perform speed estimation on cell phone data and use GPS data as ground truth.
2. Add a limitation section that discuss three main limitations of our work:
    - (M4) The lack of WiFi data in our analysis.
    - (M2b, R1-3) The length of our dataset and the limited geographic area.
    - (M2c) The limitations introduce by coarse-grained location information and irregular time intervals in the dataset.
3. (M1) Give more description on our dataset.
4. (R2-1, R4-1) **TODO: Revise the motivation of our work.**
5. (M5) Add additional references.
6. (R1-4) Add implementation details and environment settings.
7. Fix several naming and figure issues. **remaining change: figure and listing on the same page as text, fig11 and fig12 to relative value**
8. Carefully proof read to make sure there is no more language or formatting issue.

>> RESPONSE TO META-REVIEW <<

> 1. (R1) Describe how the data was collected. Who were involved? How long
> the data was collected? When the data was collected? Is it cell data the
> only collected data or was WiFi also collected?

Our dataset is collected by a major mobile carrier in China at the cell phone tower side. All active GSM data access during a three-hour period from 6pm to 9pm on a Sunday at September 2014 were recorded. Unfortunately, the dataset only contains cellular data without any WiFi data.

We mark the related change in paper as M1.

> 2. (R1 and R2) The authors should address the generality of the proposed
> algorithm as a limitation. Generally, smartphone users use both cellular
> and wifi networks in other areas. The current analysis model excludes
> many contexts of users such as location, time, individual context, etc.
> In order to understand the resulting data, it is very important to know
> about moving speed with respect to users and their culture.

a. WiFi networks. Our response to WiFi related comments is listed under meta-review comment 4 (M4) below.

b. Time and Location contexts. Due to the short time period of our dataset, it is not feasible for us to perform in-depth temporal or spatial analysis. We state this limitation in our new Section **??**.

We mark the related change in paper as M2b.

c. Individual context. We agree that our results only represent part of users who frequently moved and use mobile networks because of limitations in our dataset, i.e., coarse-grained location information and irregular time intervals. We state this limitation in our new limitation section.

We mark the related change in paper as M2c.

> 3. (R2) Speed estimation algorithm is not proved yet. The authors should
> demonstrate that their approach to estimate intra-cell movement speed is
> accurate and therefore can be reliable used for the correlation analysis.
> Many assumptions in estimating moving speed are made without much
> explanation. These assumptions should be proved and discussed in detail.
> In addition, the classes of moving speed should be analyzed and discussed
> in detail.

Since it is not feasible for us to collect ground truth of user mobility in our dataset, we use a similar but much smaller dataset to perform a brief evaluation of our speed estimation algorithm. This dataset contains both cell phone data and GPS traces. We use the GPS as ground truth for the evaluation.

We mark the related change in paper as M3.

> 4. (R1 and R2) The authors should discuss or demonstrate the implications
> of considering cellular traffic only on the results presented in this
> work. WiFi usage could be a major limitation of the generalization of the
> study.

Although WiFi handles half of the mobile data traffic and more than half of time users are connected to WiFi, cell phone network have much better coverage and user mobility diversity. It is more meaningful to study user mobility under cellular traffic. We do acknowledge that the lack of WiFi data is a limitation of our work and that our current speed estimation algorithm cannot be directly applied to WiFi data. However, incorporate WiFi data will change our speed estimation algorithm significantly and it is very hard to find WiFi data on a whole city scale. So we give possible solutions in our limitation section and are planning to solve it in our future works.

We mark the related change in paper as M4.

> 5. Suggested References:
> [1] Kyunghan Lee,Joohyun Lee,Yung Yi,Injong Rhee, and Song Chong. 2013.
> Mobile Data Offloading: How Much can WiFi Deliver? IEEE/ACM Transactions
> On Networking 21, 2 (2013), 536551.
>
> [2] Paul Baumann and Silvia Santini. 2014. How the availability of Wi-Fi
> connections influences the use of mobile devices. In Proceedings of the
> 2014 ACM International Joint Conference on Pervasive and Ubiquitous
> Computing Adjunct Publication - UbiComp âĂŹ14 Adjunct. (2014), 367372.

We have added all suggested refereces and discussed them in related works section.

We mark the related change in paper as M5.

>> RESPONSE TO REVIEWER 1 <<

> Although the authors shows interesting results, there are several
> limitations to be impored. First of all, IâĂŹm wondering how the data was
> collected. The authors mainly describes the volume of dataset. However,
> it is very important to know about how the data was collected. The
> population of the dataset might be affected by the culture and types of
> users.

Please see response to meta-review comment 1.

> Second, the proposed methodology and analysis are too general to be
> applied to understand smartphone app usage. This is because the contexts
> users use their smartphones are blurred with moving speed. Generally,
> user behaviours are divided into moving, sitting and staying at home or
> office and also related to time of day. Users spend much time on using
> smartphone when waiting or sitting for something. The usage patterns are
> also related to locations such as home or office or stores. Furthermore,
> the results could not represent the significant part of users since usage
> patterns are very diverse according to users. As the authors mentioned
> with Fig 1, the data was mainly from those who frequently moved and used.
> I think that the moving speed should be related to more contexts such as
> time and location. It might be also interesting to see the difference
> between inter-cell moving and staying within a cell.

Please see response to meta-review comment 2.

> Third, I think that the results are very limited to specific areas.
> Although the data were from three cities and the size of data is
> comparable with other countries, the results are only for the China. I
> think that the Fig 13 ~ 15 are not meaningful for understanding general
> usage patters, but for three cities of China. The authors need to
> describe the limitation of their analysis and method.

Agree. We discuss this in our new limitation section. The results may be limited to specific areas. However, the speed estimation method can be easily applied to similar datasets from other areas. Similar procedure can be used to study usage patterns in those datasets without much difficulty.

We mark the related change in paper as R1-3.

> Lastly, I would like to see how the authors dealt with the massage
> smartphone usage data.

We use Python to analyze our dataset. More specifically, we use voronoi package from Scipy to construct voronoi maps with tower coordinates. We use another Python package called shapely to calculate various geometry values in our computation of distance lower bound. All analysis are carried out on a single Cloudlab c8220 server with two 10-core 2.2GHz E5-2660 processors and 256GB memory.

We mark the related change in paper as R1-4.

>> RESPONSE TO REVIEWER 2 <<

> ## 1. Motivation for how the results of this work can be used and their
> generalization.
>
> The authors present correlations between user movement speed and other
> features such as number of applications used, traffic volume, idle time,
> etc.
> The authors motivate such findings with: âĂIJUnderstanding such
> correlations, if any, could provide useful contextual information for
> relevant and accurate app recommendation and ad delivery. For example, if
> we find out hiking hobbyists use certain apps considerably more often,
> then such apps may be more useful venues for ad delivery for equipment
> makers for hiking activities.âĂİ
>
> However, from the findings presented in this work I only see that Fig16
> supports the aforementioned motivation by providing insights about the
> correlation between movement speed and app categories.
> Providing stronger / additional arguments why all the other observations
> are relevant would help the authors to increase the value of this work.

We appreciate your thoughtful suggestions.
We mark the related change in paper as R2-1.

> Furthermore, as mentioned in the paper (however, only once I guess), the
> traces used for this work cover cellular communication only.
> As the authors have pointed out, part of the communication might be
> handled over Wi-Fi.
> In recent studies, authors have observed that over 60% of the time users
> are connected to Wi-Fi [1] and that half of the traffic is typically
> handled over Wi-Fi [2].
>
> So the important question at this point is: how representative are the
> results presented in this work given the fact that they consider cellular
> app usage and traffic only?

Thank you for your comments. Please see response to meta-review comment 4.

> ## 2. My second concern is about the performance in estimating movement
> speed that has a direct influence on the results presented in this work.
>
> To estimate intra-cell speed, the authors propose a novel algorithm that
> is based on a set of assumptions such as that users move with a constant
> speed (4.1) or with a âĂIJstraight line trajectoryâĂİ (4.3).
> The resulting computation of the movement speed is therefore a direct

> consequence of the aforementioned assumptions and the introduced
> approach.
> Therefore, the results presented in this work rely on the quality of this
> computation.
>
> The results presented in this work show a positive correlation (Fig10a)
> between movement speed and traffic volume / sec. as well as a negative
> correlation (Fig10b) between the idle time and speed.
> The conclusion is that the faster a user passes a given cell, the more
> bytes/s she will generate and the smaller the idle time intervals are.
> So to make sure that these insights indeed cover a valid collection
> between movement speed and other features, it is mandatory to show that
> the aforementioned assumptions hold, in general, and the movement speed
> estimation is accurate, to some extent, allowing to make conclusions from
> the experiments.
>
> Addressing this shortcoming might for instance include an evaluation of
> the approach on a data set that contains GPS and cellular data.
> There are several publicly available data sets that might be helpful at
> this stage (reality mining, nokia, lifemap, etc.).
> Alternatively, running a small custom study to verify the assumptions and
> get quantitative evidence that movement speed estimations are accurate,
> might also be an option.
>
> Without showing that the novel approach presented in this work to compute
> movement speed produces reliable estimates, it is at least possible that
> to some extent the insights presented in this work result from the
> inaccurate movement speed estimations.

We are really appreciate for your constructive comments. Please see response to meta-review comment 3.

> ## Minor comments:
> - Please try to put Figs and Listings on the same page as the text
> which refers to them
> - Inconsistent writing of PBE in Sec 4.1 and Sec 4.2
> - Potential naming (variables) inconsistencies between Alg1 and Sec 4.2
> - Fig10a y-axis label: is it not supposed to be âĂIJbytes/secâĂİ as
> explained in the corresponding section?
> - Sec 5.3: how do you define âĂIJa data accessâĂİ. is it a single CDR in
> the data set?
> - What is the overall value / take-home message of Fig11 and Fig12? Is
> it not better to have relative values if the number of instances differ?
> - Fig14 and Fig15 should be a bit smaller to match the font size of the
> text

Thank you for identifying these problems. We corrected these problems accordingly. We are sorry that we are not able to use red color to highlight some changes.

> Suggested References:
> [1] Kyunghan Lee,Joohyun Lee,Yung Yi,Injong Rhee, and Song Chong. 2013.
> Mobile Data Offloading: How Much can WiFi Deliver? IEEE/ACM Transactions
> On Networking 21, 2 (2013), 536551.
>
> [2] Paul Baumann and Silvia Santini. 2014. How the availability of Wi-Fi
> connections influences the use of mobile devices. In Proceedings of the
> 2014 ACM International Joint Conference on Pervasive and Ubiquitous
> Computing Adjunct Publication - UbiComp âĂŹ14 Adjunct. (2014), 367372.

Please see response to meta-review comment 5.

>> RESPONSE TO REVIEWER 4 <<

> the use case presented (that of targeted advertising) needs evidencing
> or reframing. There are many ways to target user interests and it is not
> obvious that an approach based on speed is appropriate. If there is no
> evidence to support the use of speed in targeted adverts then I would
> have preferred to have seen the approach based as a generic analysis with
> a range of possible applications being suggested.

We appreciate your thoughtful suggestions.
We mark the related change in paper as R4-1.

> clearly the data captured is only part of a user's overall data
> consumption. This point is made in the paper but should be made more
> explicitly at the start.

Please see response to meta-review comment 1.

> the length of the data trace (3 hours) seems short - it would be good
> to explain why the authors feel this is an appropriate amount of data to
> analyse.

Thank you for your comments. Unfortunately, this is all the data that available to us. We have discussed this limitation of our dataset in the new limitation section. Eventhough our dataset is relatively short, the speed estimation method can be easily applied to similar datasets of any given time period. Similar procedure can be used to study usage patterns in those datasets without much difficulty.
We mark the related change in paper as M2b.

> the speed estimation approach contains a very large number of
> assumptions and there is no evidence presented that it actually works.
> This is the biggest weakness in the paper - it *must* provide some
> evidence of success. Even a simple study with 20 users where ground truth
> and cell records were collected would be sufficient. Without that there
> is simply not enough to convince the reader that the algorithm presented
> works.

Please see response to meta-review comment 3.

> the analysis would be much stronger if it was backed up with some
> evidence (e.g. a survey or a focus group or some observations studies)
> that trued to explain the patterns seen.

Thank you for your comments. However, we are not able to find previous survey or observation studies on this topic.