



> \*\*\* Detailed Comments: Please provide detailed comments that will help the TPC assess the paper and help provide feedback to the authors.

I liked the work. There are some issues that should be addressed, some may be resolved through better clarifications, others (eg the cross validation) may need some work.

- Effect of WIFI

Your user analysis results are not absolute as they do not account for the presence of WiFi hotspots which are likely to bias the results that you are seeing. At work and at home people who have WiFi tend to use WiFi. The top locations you see for a user may be a function of the WiFi availability for that user. Both the amount of time and amount of data one uses at a location will be influenced by if one has WiFi connectivity there.

Make it clear in the paper, that your analysis is conditioned on WiFi deployment.

Instead of ramping up the bandwidth in the mobility network, another option is to consider selectively deploying WiFi hotspots. Many providers are taking this route. This is attractive since often the most constrained resource is the last mile spectrum in mobility networks. How would that compare to your proposal in terms of cost and effectiveness?

- How do you get the base station location of the user from the billing trace? How accurate is it?

As I understand, the base station location in these types of billing systems is obtained from tunnel status messages for the GTP tunnel between the SGSN and GGSN over which a handset's data connection runs. The base station info is correct when the GTP session is initiated. But afterwards it can get stale if the user is moving. The GTP update messages only update the base station info for handovers that involve SGSN handovers. For handovers between base stations covered by the same SGSN, the GTP info is not updated. A single SGSN typically covers many base stations over a large area. So your base station information can be quite stale, particularly since GTP tunnels and states can last for the order of many days.

It's possible that the authors' trace has augmented information from other sources to rectify the above problem. The

authors should explain how the base station information was populated in their trace, how accurate it is, if they validated the data quality of the base station info they use and quantify how inaccuracies in this data impact their findings.

- Evaluations: The DropZone placement is formulated as a static optimization problem. I had expected to see some validation experiments where the placement is done using one part of the data and evaluations using other data. I did not see that. This is pretty standard and required to understand what happens when the user behavior changes from the one it was optimized for.

> \*\*\* Recommendation: Your overall rating (Please try giving as few borderlines as possible).

B+ = (top 20% of reviewer's perception of all INFOCOM submissions, but not top 10%) (4)



many total base stations exist in the data set studied, and the percentage of drop zones selected from this pool. This will make some analysis much clearer as opposed to using the absolute # of drop zones, as it currently shows in fig 6 and 9.

> \*\*\* Recommendation: Your overall rating (Please try giving as few borderlines as possible).  
B+ = (top 20% of reviewer's perception of all INFOCOM submissions, but not top 10%) (4)

===== Review 3 =====

> \*\*\* Contributions: What are the major issues addressed in the paper? Do you consider them important? Comment on the novelty, creativity, impact, and technical depth in the paper.

The paper relies on some real mobility traces of the users of a cellular network, to characterize their data uploading process. The conclusions of this study is that each user uploads most of the content by a limited number of locations and that she/he is usually postponing the upload in comparison to the time the content has been generated. These observations motivate their proposed solution to select only some specific locations (drop zones) to upgrade the cellular infrastructure and provide incentives to let the users postpone their uploads until they do not reach these zones.

A heuristic is proposed to determine the position of such locations and experiments on the same traces are carried out to show the number of dropzones needed to be able to upload through the drop zones a given percentage of the total traffic under a constraint on the maximum time the upload is postponed. The conclusions from the analysis of the traces are interesting, even if I have some remarks about their validity. To the best of my knowledge the proposed solution is new, even if it is not a breakthrough. It is not clear why the drop zones could not be simply WiFi access points. Performance evaluation would benefit from a comparison with the performance of the existing cellular network.

> \*\*\* Strengths: What are the major reasons to accept the paper? [Be brief.]

The change of usage traffic patterns due to the spread of smartphones is a real challenge for user operator.

The analysis of real traces is the key element in this paper.

> \*\*\* Weaknesses: What are the major reasons NOT to accept the paper? [Be brief.]

The conclusion about users being willing to postpone their uploads is questionable.

It is not clear why users could not be incentivized to use alternative access networks like WiFi, or wired networks.

> \*\*\* Detailed Comments: Please provide detailed comments that will help the TPC assess the paper and help provide feedback to the authors.

I first explain my remarks about the main weaknesses of the paper.

1) The conclusion about users being willing to postpone their uploads is questionable.

The authors show that users often postpone their uploads after many hours from the generation of the content. From this observation they conclude that users may be willing to postpone their uploads and this would support the their drop zones solution.

In reality, it seems reasonably that the upload occurs at a later time for one of the following reasons:

- i) the users have not decided yet to upload the content,
- ii) they do not have time to upload it at the generation time (they are moving and they were just able to took the photo),
- iii) they have experienced bad performance when uploading while moving and so they postpone their upload to a later moment when they are at one of their comfort zones.

These reasons are in agreement with the fact that they upload from a limited number of locations. At home or at work they have the time to process the content generated and decide what to do with it, they have the time to launch the upload, and they may experience better uploading performance. Then, in my opinion, the observation of late uploads does not provide evidence that users may be willing to postpone them.

2) It is not clear why users could not be incentivized to use alternative access networks like WiFi, or wired networks.

The authors think that a software should be installed on the smartphones to ask the users if they are willing to postpone their upload and to transparently do it when the user reaches a drop zone. The authors think that users should be incentivized to choose this option. Now it seems that these incentives could also promote the use of other cheaper networks that could be available later, like WiFi connections (most smartphones support WiFi), or wired connections.

If users can really be convinced to postpone their upload, why would it be necessary to deploy smart zones rather than taking advantage of existing alternatives?

#### OTHER REMARKS

- A comparison with the current cellular architecture is missing.

Figure 9 motivates the conclusion that the deployment of 963 base stations with LTE technology could be able to handle a 4 order magnitude increase of the traffic (expected to happen in 14 years) while still managing 50% of the content upload through drop zones. I would have appreciated a comparison with the current cellular network: what percentage of traffic increase could it handle if no upgrade was performed?

- Figure 8.b about the average distance from a drop zone

As the authors state, the increase on the average distance is counter-intuitive and I have not understood their explanation. It is possible that the authors are not showing the average distance of a user from the closest drop zone to her/him (then the average is evaluated across the users), but rather the average distance of users from all the drop zones (then the average is evaluated across the users and for each users across the drop zones). This would justify the result. Please clarify.

> \*\*\* Recommendation: Your overall rating (Please try giving as few borderlines as possible).

B+ = (top 20% of reviewer's perception of all INFOCOM submissions, but not top 10%) (4)