How the Availability of Wi-Fi Connections Influences the Use of Mobile Devices

Paul Baumann
Wireless Sensor Networks Lab
TU Darmstadt, Germany
paul.baumann@wsn.tu-darmstadt.de

Silvia Santini
Wireless Sensor Networks Lab
TU Darmstadt, Germany
santinis@wsn.tu-darmstadt.de

UbiComp '14, September 13 – 17 2014, Seattle, WA, USA

ACM 978-1-4503-3047-3/14/09...$15.00.
http://dx.doi.org/10.1145/2638728.2641703

Abstract
Several aspects might influence the way users operate their mobile devices in a mobile context. In this work, we show how the presence or absence of a Wi-Fi connection influences the amount of data traffic generated by mobile devices. Our results show that the probability of users to generate data traffic while connected to Wi-Fi is twice as high as when a cellular connection only is available. Furthermore, we observe that an almost constant amount of data traffic is generated over a day, although it slightly increases in the late afternoon. Last but not least, we observe that fair-use policies do not seem to influence the behavior of mobile users with respect to the amount of traffic they generate over different weeks of a month. We ground our analysis on the Device Analyzers data set, which contains detailed records of mobile phone usage of more than 17,000 users from all over the world. Building upon these preliminary quantitative results, we outline how the availability of data for mobile users can be improved by combining mobility and phone usage prediction with knowledge about the temporal and spatial availability of Wi-Fi connections.

Author Keywords
Mobile device usage, data set analysis

ACM Classification Keywords
H.m [Information systems]: Miscellaneous.

Research Question, Approach, Related Work, and Key Findings

The widespread availability of mobile devices and data connections makes users increasingly rely on them to carry on daily activities. Examples of such activities include setting location-based reminders about buying specific items, accessing documents from everywhere, or enjoying the latest episodes of a TV series while commuting to or from work. The possibility to generate – i.e. send and receive – data traffic depends however on the availability of a corresponding data connection. Many mobile phone users can rely nowadays on flat-rate plans offered by mobile operators, which provide users with "unlimited" access to the Internet from their mobile devices. Flat-rate plans however often do limit the amount of data traffic users can generate within a given period of time. Also, the coverage and bandwidth provided by mobile phone operators' data connections might not always be sufficient to guarantee a desired level of user satisfaction. The availability of Wi-Fi access points might compensate for these limitations by providing a fast – and often free! – data connection. This might be enabled by open Wi-Fi access points or access points made available by employers to their employees.

In this work, we show that the availability of different types of data connections influence the behavior of mobile device users with respect to the amount of data traffic they generate. In particular, we show that users tend to actively use their phones and generate data traffic with a higher probability when they can rely on a Wi-Fi connection rather than when they are connected to the cellular network. We also observe that users tend to generate the same amount of data traffic over the day, although this amount is slightly higher in the late afternoon. Further, our results show that fair-use policies do not seem to influence the amount of data generated by users over subsequent weeks of a month.[1] We obtain these results by analyzing the mobile phone usage statistics of 790 users whose data records are included in the Device Analyzer data set [3].

The ultimate goal of performing this type of quantitative analysis is to provide means to pre-load data on mobile devices depending on both the current and future availability of Wi-Fi connections as well as on user mobile usage habits. The availability of Wi-Fi is indeed typically bound to specific places and users show high regularity in their mobility patterns [1, 7]. Users tend to spend about 80-85% of their time at home, in the office or other "indoor" places and only about 15-20% of their time in a mobile context, e.g., travelling from one place to another [8]. Church *et al.* [6] and Teevan *et al.* [5] have studied users' search behavior on mobile devices for a number of situations including the mobile context. They show that despite the fact that users tend to spend only about 15% of their time in a mobile context, about 67% of their information needs – and thus of the data traffic they generate – arise during this time. Given this high interest of accessing data in a mobile context, our focus is on getting a deeper understanding of which data is relevant for users right before they start moving from places that provide a Wi-Fi connection to places that do not. This for instance happens when users commute from home to work or vice versa. Commuters often tend to read news or watch videos while in transit. Data users access in a mobile context could conveniently be loaded on the mobile device before leaving places with a Wi-Fi connection thus avoiding the need of relying on the data connection of mobile phone operators.

[1] A fair-use policy is a strategy that allows mobile phone providers to limit the bandwidth available to a user if the amount of data traffic she generates in a given time period exceeds a pre-specified amount (e.g., 1 Gigabyte over 1 month).

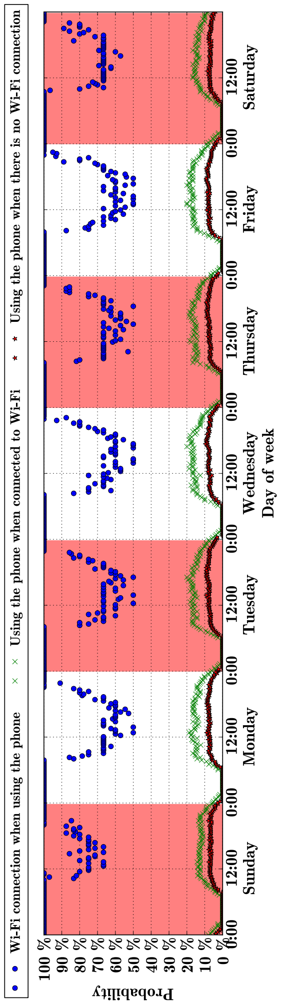

Figure 1: Median probabilities aggregated for a week over a observation period of multiple months and 790 users.

The results presented in the next section represent a first step towards the design of techniques for data pre-fetching that combine both mobility and data usage prediction.

Analysis of the Device Analyzer Data

To investigate the research questions outlined above we leverage the Device Analyzer data set [3]. This data set has been made publicly available for the first time in the context of the UbiComp/ISWC 2014 Programming Competition.[2] The released data contains records of mobile phone usage of more than 17,000 users across the world. For our analysis we selected users for which data collected over a period of at least 10 months is available. Thus, the results described below are based on data from the 790 users that satisfy this criterion. The specific sensor data we consider includes records of battery level, screen status, last accessed application, and network statistics information.

In the following, we present three selected results from our analysis. The first result shows that users tend to actively use their phones and generate data traffic with a higher probability when they can rely on a Wi-Fi connection rather than when they are connected to the cellular network. The second result shows that users tend to generate the same amount of data traffic over the day, although it is slightly higher in the late afternoon. The third result shows that fair-use policies do not seem to influence the amount of data generated by users over subsequent weeks of a month.

The Availability of a Wi-Fi Connection Fosters the Use of Mobile Devices

To derive the first of the three results mentioned above we consider time slots of length T_{slot}=15 minutes, which result in a day being divided into 96 slots. For each day of the week and each time slot we then count the number of time slots in which a user has generated data traffic while actively using the phone. We consider a user to be actively using the phone if the screen of the phone is on (this information is included in the data set). We also differentiate between data traffic being generated when the phone accesses the Internet through a Wi-Fi connection or using a cellular data connection (e.g., GSM or LTE) provided by the mobile operator.

We thus compute for each slot s and user i the values of the following counters: $N_{use}(s,i)$ is the number of times user i has generated data traffic in slot s while actively using the phone; $N_{use,wifi}(s,i)$ is the number of times user i has generated data traffic in slot s while actively using the phone and while the phone was connected to the Internet through Wi-Fi; $N_{wifi}(s,i)$ is the number of times the phone of user i was connected to the Internet through Wi-Fi in slot s; $N_{cell}(s,i)$ is the number of times the phone of user i was connected to the Internet through a cellular network in slot s.

The data points plotted through round markers in Figure 1 show the median value, for each slot and computed over all users, of the ratio $N_{use,wifi}/N_{use}$. In other words, each data point shows the median relative number of times (or: the median empirical probability) of a user being connected to Wi-Fi while being generating data traffic and actively using the phone. The plot shows that during the day in more than 50% of the instances the user is connected through Wi-Fi when she generates data traffic and actively uses the phone. In the rest of the instances the user relies on the connection provided by the mobile phone operator. The plot also shows that the (median) probability of the user to exploit a Wi-Fi connection during the night is 100%. This is most likely because most users have an own Wi-Fi access point at home, where they also spend most of their nights.

[2] http://ubicomp.org/ubicomp2014/calls/competition.php

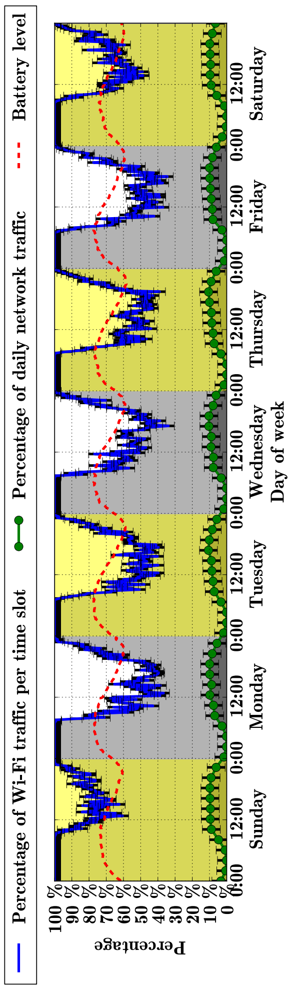

Figure 2: Median percentage values aggregated for a week over a observation period of multiple months and 790 users.

The data points plotted using crosses (x) as markers show for each slot the median value, computed over all users, as the ratio $N_{use,wifi}/N_{wifi}$. I.e., these data points show the probability of a user to actively use the phone and generate data traffic when she is connected to the Internet through Wi-Fi. The data points plotted using stars (*) as markers show for each slot the median value, computed over all users, of the ratio N_{use}/N_{cell}. I.e., these data points show the probability of a user to actively use the phone and generate data traffic when she is connected to the Internet through the cellular network. The systematic difference between the two curves shows that users are about twice more likely to actively use the phone and generate data traffic when a Wi-Fi connection is available rather than when only the cellular network is available.

Overall Generated Data Traffic

We now consider the amount of generated data traffic for each time slot of a week. To this end, we sum the generated data traffic over the Wi-Fi and cellular network for each user and each time slot. We then compute the percentage of traffic that has been generated over Wi-Fi, plotted as a median value across all users in Figure 2 by the solid line. The black bars cover the 95% confidence interval. We observe that during the night, the traffic is mainly generated over Wi-Fi, since most of the persons are at home. During the day, the percentage of the generated data traffic that goes over a Wi-Fi connection drops below 50%. The shape of the plot has also a strong correlation to the plot in Figure 1 that represents the probability of having a Wi-Fi connection when accessing the network data.

Furthermore, Figure 2 shows the percentage of daily transmitted network traffic per 2-hours time slots, resulting in 12 time slots per day. We observe a constant shape over the whole day with a small peak during the evening hours and minimum values during the night. Considering these results together with the results for the probability of having a Wi-Fi connection for each time slot, we observe a negative correlation. For instance, the peaks of the transmitted data are at the same time periods as the lowest values for the probability of having a Wi-Fi connection.

Influence of the Fair-Use Policy on the Transmission of Network Data

Finally, we present the results of a preliminary analysis aimed at verifying whether fair-use policies applied by mobile phone operators influence the amount of traffic generated by users. For each user we first compute the average data traffic relayed both through a Wi-Fi or cellular connection over the first four weeks of a month. We omit the last week of each month since it only covers at most three days (months with 31 days). The traffic for the different months but the same number of weeks in each month are combined together. This process provide us with four values for each user. One value per week that indicates the average transmitted network traffic. For each user we then normalize these four values by the amount of data of a week in which most of the data has been transmitted. For instance, if a user transmits 1MB, 2MB, 1MB, and 4MB on average in each week of a month then the resulting values are 0.25, 0.5, 0.25, and 1 for each transmitted amount of data respectively. We refer to these values as the *relative fraction* of the generated data traffic. This step allows us dealing with the different amount of recorded days across all the users. In our first analysis, we assume that the mobile phone operators restore the users' traffic counter at the beginning of each month. In our future work, we will consider defining this event more precisely.

Given this process, Figure 3 shows the relative fraction of the generated data traffic for the first for weeks of a month.

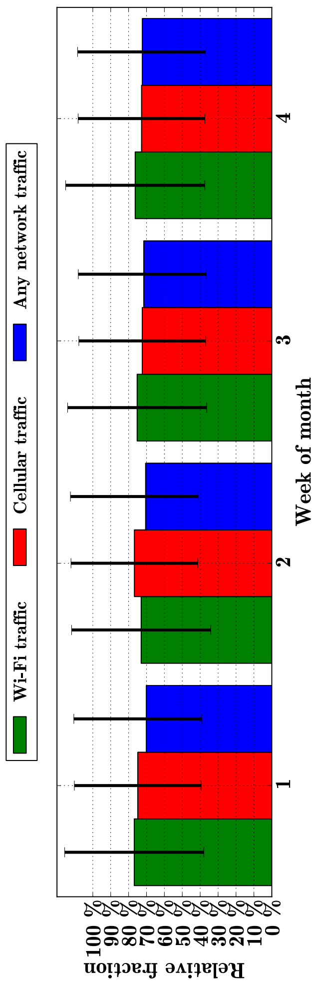

Figure 3: Relative fraction of the average transmitted network traffic shows as a median values of 790 users. Whiskers cover 95% confidence interval.

Each bar indicates the median value across all users. The corresponding error plot covers the 95% confidence interval. We observe that there is no decrease in cellular traffic consumption over a month. This might have been expected if the users were aware of the already transmitted cellular traffic and the remaining limit before the fair-use policy will be enforced. We have run the same study for the top 5% of users that transmit on average most of the traffic over the cellular network. The results are very similar to Figure 3. We conclude that our first analysis on the potential influence of the fair-use policy does not provide any evidence that the policy influence users' usage of the cellular network data. However, one potential explanation, which we will verify as part of the future work, is that the fair-use policy is reset on different days of a month. Given this hypothesis, we aim to find days at which the generated data traffic over the cellular network breaks down.

Conclusions and Future Work

In this work, we have shown that there exist several factors that prevent users from using their phones as preferred while they are in a mobile context. In particular, we have focussed on the role of the availability of Wi-Fi connections. Our results reveal that users tend to actively use their phones and generate data traffic with a two times higher probability when they can rely on a Wi-Fi connection rather than when they are connected to the cellular network.

In our future work we plan to further leverage the Device Analyzer data set [3] to better understand the reasons behind this behavior. To this end, we plan to use data from users that have released non-anonymized data about the applications they used on their phone. Furthermore, we plan to combine this information with data about the places usually visited by users. This is motivated by the fact that a high number of information needs arise in a combination with a current spatial context [9]. The users' places will be extracted by leveraging the anonymized Wi-Fi data and one the state-of-the-art algorithms [2, 7].

Furthermore, given the presented results we will focus on the fraction of users that have shared the names of their applications to understand the influence and the possible benefits for users by pre-fetching the data. The idea is to connect the ability to pre-fetch applications' network data while the mobile device still experiences a Wi-Fi connection with the ability to predict human mobility. For instance, by predicting how long the user will stay in the vicinity of a connected Wi-Fi network, which place she will visit next, and the understanding of which information needs in terms of accessed applications will be required, we aim to pre-fetch this information in advance. Teevan *et al.* [5] have demonstrated that users' information needs are highly contextual. Furthermore, Yan *et al.* [9] have shown a correlation between the used applications and the current visited place as well as the temporal aspects, e.g., time of day.

Building upon this idea, we will differentiate between the network data that can be pre-fetched, e.g., latest TV show episode, and live-generated content, e.g., messaging. By leveraging the Device Analyzer data set we will be able to estimate the potential cellular traffic savings. Falaki *et al.* [4] have considered the "interactive" traffic, i.e., traffic that is generated while the user was interacting with the phone or at least the screen was on. Their results indicate that most of the traffic is generated in these phases. However, the definition of "interactive" traffic does not allow understanding which traffic can be pre-fetched in advance and which traffic cannot, e.g., communicating with friends. In order to analyze whether this concept allows users using their mobile device in the way how they would prefer to do it in a mobile context, we will run a dedicated study.

Beside the influence of the absence of a Wi-Fi network connection, we will consider further potential sources that might prevent users from using their mobile devices as they prefer it in a mobile context. One of these sources might be the mobile devices' battery level. By leveraging the battery level information in the Device Analyzer data set, we computed the users' average battery power for each weekly 15-minutes time slot. Figure 2 shows the median values of the average battery level for each time slot across all the considered users as a dotted line.

We observe multiple insights. The curve exhibits a waveform shape with peaks in the morning hours and minimums in the evening hours. Furthermore, these points correlate with the time slots at which the median probability of having a Wi-Fi connection while accessing the network data is for the last time at 100% in the morning of each day. At the same time, the minimum battery level correlates with the time slots at which the probability of having a Wi-Fi connection becomes 100% again. We expect that these points mark the situations at which users usually leave or arrive at home. By considering the spatial data and the information about the charging status in the Device Analyzer data set, we expect to confirm this statement.

We further observe that the median value of batteries' average level is with 60% at the lowest point. However, our hypothesis is that in situations in which users expect to not be able to recharge their mobile device for a long period of a day, they adapt their appropriate usage behavior of their mobile device. By leveraging the Device Analyzer data set, we plan to confirm or decline this hypothesis and provide potential solutions for the caused implications.

Acknowledgements

This work has been partially supported by the Collaborative Research Center 1053 funded by the German Research Foundation and by the Priority Program Cocoon funded by the LOEWE research initiative of the state of Hesse, Germany.

References

[1] C. Song *et al.* Limits of Predictability in Human Mobility. *Science 327*, 5968 (2010), 1018–1021.

[2] D. Kim *et al.* Discovering Semantically Meaningful Places from Pervasive RF-Beacons. In *UbiComp* (2009).

[3] D. Wagner *et al.* Device Analyzer: Understanding Smartphone Usage. In *MOBIQUITOUS* (2013).

[4] H. Falaki *et al.* Diversity in Smartphone Usage. *MobiSys* (2010).

[5] J. Teevan *et al.* Understanding the Importance of Location, Time, and People in Mobile Local Search Behavior. *MobileHCI* (2011).

[6] K. Church *et al.* Understanding the Intent Behind Mobile Information Needs. *IUI* (2009).

[7] P. Baumann *et al.* The Influence of Temporal and Spatial Features on the Performance of Next-place Prediction Algorithms. In *UbiComp* (2013).

[8] S. Brasche and W. Bischof. Daily Time Spent Indoors in German Homes–Baseline Data for the Assessment of Indoor Exposure of German Occupants. *Intl. Journal of Hygiene and Environmental Health 208*, 4 (Jan. 2005), 247–253.

[9] T. Yan *et al.* Fast App Launching for Mobile Devices Using Predictive User Context. *MobiSys* (2012).