Deriving Connectivity and Application Usage Patterns From Longitudinal Mobile Phone Usage Data

Sokratis Barmpounakis
Institute of Services Science
University of Geneva
Carouge, Switzerland
sokratis.barmpounakis@unige.ch

Katarzyna Wac
Institute of Services Science
University of Geneva
Carouge, Switzerland
katarzyna.wac@unige.ch

ABSTRACT

In this paper, based on the Mobile Data Challenge data obtained from the Lausanne data collection campaign, our research aims first to derive network connectivity (e.g., WLAN, 3G) and its Quality of Service (QoS) and mobility patterns of the mobile users, as this connectivity and QoS relate to the user's application activity. Second, we aim to understand how these patterns relate to the overall Quality of Experience (QoE) of the user. Concerning the mobility patterns, we define indoor and outdoor activity for each mobile user. Moreover, we attempt to define semantic places using time filters and GIS techniques, which could also be correlated to the application activity of the users. By correlating the above with the application activity of the users, as well as the hour and weekday patterns, certain inferences can be extracted, concerning the users' spatial and temporal behaviour. These inferences could be used further in developing methods for assurance of the mobile users' QoE.

Categories and Subject Descriptors

C.3.3 [**Performance of Systems**]: Reliability, availability, and serviceability

General Terms

Algorithms, Design, Experimentation, Human Factors

Keywords

Mobile phone application usage, connectivity, mobility, location, patterns extraction, longitudinal data, data mining

1 INTRODUCTION

Wireless networks are already evolving into the basic component of communication and information transmission technologies, as well as one of the most important factors affecting the experience of a smartphone user. Common daily routines of such a user involve several connections and disconnections from different types of wireless networks. While connected to these networks, the user attempts to take advantage of what the connection offers, in a variety of ways by browsing the Web, using VoIP, watching streaming video, or, in general, using networked multimedia applications. Quality of the network service (QoS) and the quality of these applications (or smartphone activity in general) are interrelated, as there are different performance requirements for different types of applications [1].

There is no fixed pattern or rule for the majority of users indicating what kind of applications they will be using depending on the network they are connected to, or their location state. Actually, the research shows that the users' activities and interactions with the device are very diverse [3]. However, analysing the data from specific aspects like temporal or spatial activity and connectivity, useful patterns can be extracted, leading towards innovative methods for assurance of the Quality of Experience (QoE).

In this paper, we present results acquired in Mobile Data Challenge[1] [4], based on analysis of data acquired from the Lausanne data collection campaign [5].

This paper is structured as follows. Section 2 describes the way we have analyzed data in our approach. Section 3 presents the results of our analysis, while Section 4 discusses the assumptions we have made throughout our analysis. Section 5 summarizes the conducted data inferences and sketches the future work areas.

2 METHODS

2.1 General Description

The overall goal of our analysis was to derive user's wireless access network connectivity, as well as application usage patterns. Therefore, the MDC data on which we focused were the *application* logs, as well as the data enabling us to derive user's location and his/her network connectivity conditions, hence: the *gsm*, the *gps* and the *wlan_loc* log files. In order to correlate and integrate the different types of data, we used the *timestamps* data item, assuming the logs were synchronized.

2.2 Applications Used

Our analysis focuses on discovering patterns concerning the user's wireless access network connectivity, thus, we are interested only in applications, which are network-dependent and make use of online application data exchange between a mobile node and a remote server [2]. Moreover, to manage discrepancy in applications stemming from different phone language settings (*e.g.*, Maps (EN) being the same as Cartes (FR)), we have used the unique application UID number to identify each used application.

The final outcome was 7 different application clusters: *Web*, *E-mail*, *Maps*, *VoIP*, *Search*, *MMS* and *Online Sharing*. These 7

[1] This material was prepared for the Mobile Data Challenge 2012 (by Nokia) Workshop; June 18-19, 2012; Newcastle, UK. The copyright belongs to the authors of this paper

clusters were chosen with the goal to create distinctive types of interactive mobile applications, which are characterized by different traffic models. *Web* is a real-time interactive application, with purely online web-services based functionality. *E-mail* is non-real time, online application, where the initial delay of synchronizing new messages plays a major role in the user experience. *Maps*'s functionality is mainly real-time, online, however, there is the possibility for the user to change to offline functionality [6]. *VoIP* refers to "*Fringe*" and "*EasyVoIP*" applications, which were used by the participants. *VoIP* is labelled as real-time, online type of application. Yet, depending on its function (e.g., sending file, voice call), it can be online, real-time or even non-real time. *Search* is an online real-time application type. *MMS* as well as *Online sharing*, finally, refer to online, non-real time applications. It should be noted that E-mail, SMS and MMS were labelled under "*Messaging*" application, being an "envelope" Symbian OS-based application [7].

In the *application.csv* logs there are different kind of events available: *Application.started*, *.view*, *.foreground* and *.closed*. We used the *Application.foreground* event as evidence for the user's usage of this application at a particular moment, supported by the *closed* event, for retrieving the application name.

2.3 Network Connectivity

We assume that the network type the user is connected to, can be one of the 3 types: WLAN, 2G (+) or 3G (+). By 2G(+) we refer to the second-generation wireless telephone technology, plus the evolution of 2G technology: GPRS and EDGE (usually referred to as 2.5 and 2.75G). In the 3G(+) we cluster UMTS, CDMA (mainly used in the US), as well as the enhanced 3G type: HSPA(+). In the case of cell network type used at a given time instance, the distinction is based on the value of its *Cell ID* (integer). If the *Cell ID* is lower than the *65535*, we assume the user is connected to a 2G(+) network [8], because for GSM networks the cell identity is 16-bit number (2 B). If the cell ID is equal or higher to 65535, the user is connected to a 3G(+) type of cell network.

As far as the WLAN connectivity is concerned, from the MDC data description we conclude that users were using by default WLAN network when available, as their mobile providers' data plans were mainly of a pre-paid type; being expensive hence discouraging the users to use data over 2G or 3G network on a frequent basis. Following this, we make the assumption that whenever there is a trace in the WLAN_LOC table, the user is actually connected to the particular Access Point (AP), and not just being in its coverage. Moreover, based on the fact that the WLAN_LOC sampling frequency was 120 seconds, each WLAN trace found, "locks" the connectivity flag as *WLAN* for the next 120 seconds, after which the algorithm re-starts analysis of the data to update, if needed, the network to which the user is connected.

2.4 Location

2.4.1 Indoor, Outdoor and "Non-indoor" States

For the location part of the application user, we make the following assumptions: when there are *GSM* traces but no *GPS* available, we assume that the user is indoors. When there are *GPS* traces available, the user is not indoors. The expected frequency of the GPS measurements is equal to 10 seconds, and when it is evidenced in the data, we label the data as *outdoors*. If the frequency of the GPS is lower, *e.g.*, when the GPS signal perception may be weak - driving inside his car, or when next to high buildings in the city center (referred to as "urban canyon"), we label the state as *non-indoors.* Therefore, the outdoors and non-indoors states include outdoors mobile and fixed locations of the user.

2.4.2 Defining Semantic Places

Using *time* filters, some of the most significant semantic locations of the users were defined. Based on "common sense" assumptions (highest chances for a person to be at given location), it was possible to filter the location traces (*GSM*, *GPS*, *WLAN_LOC*) in a way, that certain semantic places could be discovered. For instance, applying a filter for *Monday to Thursday, from 02:00 to 06:00*, we have defined the semantic location *home*. In the same way *Monday to Friday* (weekdays), *from 10:00 to 18:00* defined *work* (or school, university, *etc.*), and *18:00 to 23:00* for weekdays defined *after work* semantic places. In combination with GIS techniques (*e.g.*, Point Density tool for *GPS*, *GSM* or *WLAN_LOC* points, *c.f.*, Figure 1) we were able to confirm the semantic locations. For ach given location we considered a set of surrounding network CellIDs. It should be noted that in order to "translate" the different cell IDs into map coordinates, and be able to compare it with real locations on the map, we used the (non-public) *glm/mmap* API from Google [8].

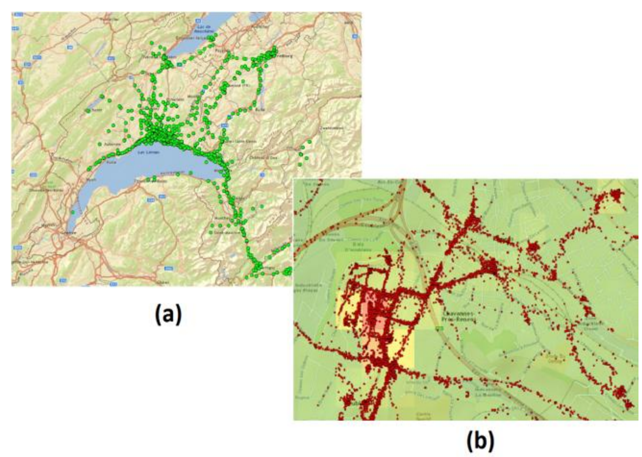


Figure 1. P002: a) GSM cell IDs' Traces Mapped On GPS Coordinates, b) GPS Density Clusters.

2.5 Main Algorithm

We combine the variables derived, as described in Sections 2.2, 2.3 and 2.4, and we correlate the application activity with the user connectivity each moment, as well as with his/her location.

The basic idea is that we scan down the unified by timestamp MDC data row by row, while handling 2 flags: one for the network and one for the location state. The algorithm's output results in instances which inform us that for example user P023, on Monday 11:00 AM, used E-mail over WLAN while indoors.

3 RESULTS

3.1 Collected Data Summary

Overall we analysed the data of 38 participants. Among them, 20 were male, 8 were female, and the gender data for the rest of the participants was missing. The age of the participants varied from below to 16, to above 50 years old. About half of them were full-time workers, 6 of them were students, while the rest were part-time workers, housewives /homemakers, retired *etc*. The subject

ID is derived from the MDC data and it does not follow the numbered ordering, *i.e.*, a subject ID ranges from 002 to 185.

The collection of the data was heterogeneous among participants. It started on September 30th, 2009 (starting with 8 participants) and ended on March 31st, 2011. The participation for each user varied from 92 to 532 days, with the average participation length of ~280 days (*i.e.*, ~9 months). Moreover, there was a lot of missing data, varying from user to user. Based on the *GSM* logs and the GSM timestamp frequency, which on the contrary to WLAN and GPS does not depend on any factor like presence of WLAN AP or being outdoors, we discover that the missing data can vary from 7% (*P063*) to 84% (*P010*).

3.2 Applications Overview

Overall, participants used many different kinds of applications ("online" and "offline"), however for the majority of them, there is a common set of the top used ones. For instance, *Text Message* application belongs in the "Top 10" of application usage for 21 different users, *Web* for 31 different users, *Calendar* for 10 users. The defined earlier 7 online application clusters are popular in most of the users' activity.

3.3 Connectivity and Application Activity

As mentioned above, we correlated the application activity of each user to the type of wireless network the user was connected at the moment of the activity. Specific patterns concerning the connectivity of the users, while using -each one of the 7 analyzed application clusters, are extracted. The relative time spend in different networks (2G, 3G or WLAN) for overall application activity per each user is presented in Figure 2. The last column of the figure represents the average user's connectivity; on average mobile applications user was connected 19% of the time over 2G(+), 51% over 3G(+) and 30% over WiFi, which results on average in 70% of connectivity over cellular network.

Figure 2. Connectivity Type For Overall Application Activity

Figure 2 shows that the overall application activity for WLAN network can really vary between users as it ranges from around 10% (with the exception of *P141* who almost did not use any WLAN) up to more than 50%. The cell network (2G(+) and 3G(+) types), accordingly, ranges from around 40% to 90% of the overall application usage.

The pie charts (Figure 3) represent the application activity for each one of the examined 7 application clusters per a wireless access network type.

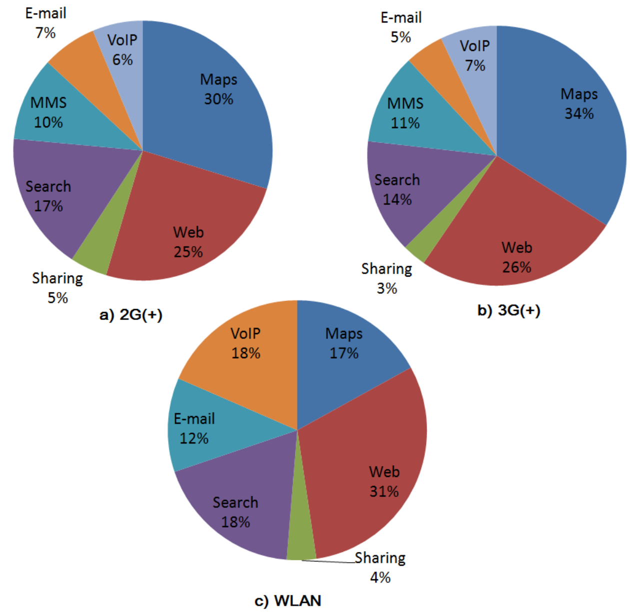


Figure 3. Application Type Per Connectivity Type (All Users)

Application clusters like *Web* and *Search* do not vary significantly with the network type. Some other, however, are handled differently in the connectivity context of the users. Two examples are *Maps*, MMS and *VoIP*. *Maps* is rarely used under WLAN network (outdoors location state may play a very important role there), MMS is impossible to be used under WLAN network due to the nature of the service, where as *VoIP* is mostly used under the WLAN network (by much less users though), possibly due to its particular network performance requirements. Figures 4 and 5 visualize use of *Maps* and *VoIP* by users under different network connectivity context. As we conclude from figures, all users but P82 used Maps at any point of time in the study, while VoIP application was used only by the self-selected 9 participants. Average user was using Maps 21% of time over 2G(+), 66% over 3G(+) and 13% over WiFi, while VoIP 14% of time over 2G(+), 42% over 3G(+) and 44% over WiFi.

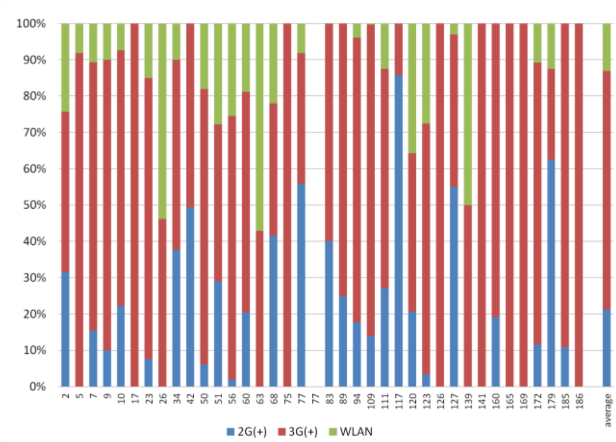


Figure 4. *Maps* Application Usage Per Connectivity Type

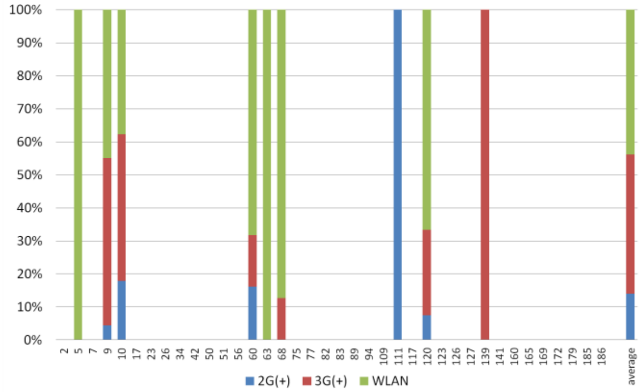


Figure 5. *VoIP* Application Usage Per Connectivity Type

3.4 Location States and Application Activity

In this section, we present results of analysis of the percentage of applications used per a location status (*indoors*, *outdoors*, "*non-indoors*"). Figure 6 shows that for all users, the largest part of the application activity is taking place indoors. On average, for all applications analyzed, the activity is *outdoors* (or *not-indoors*) in around 3% of the cases. This can be explained by the fact that in general, the time spent by users is mainly indoors. The conclusion that mobile applications are mainly used indoors was also reached in our other user studies as reported in [10].

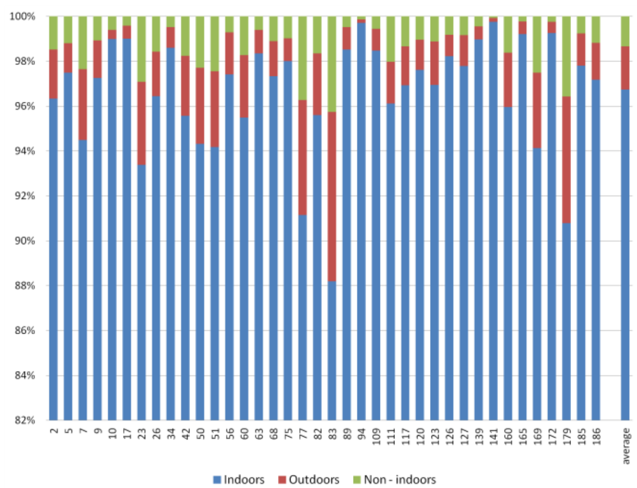


Figure 6. Location Type For Overall Application Activity

The distribution, however, of the indoor and outdoor application activity per application, proves that users make different use of applications when indoors, or outdoors (possibly moving). For example, one of the users - with relatively "typical" application activity patterns - *P009*, makes higher use of *Maps* and *Search* applications, when outdoors (Figure 7), than the *outdoors* application activity as average indicator, which is less than 5% (Figure 8).

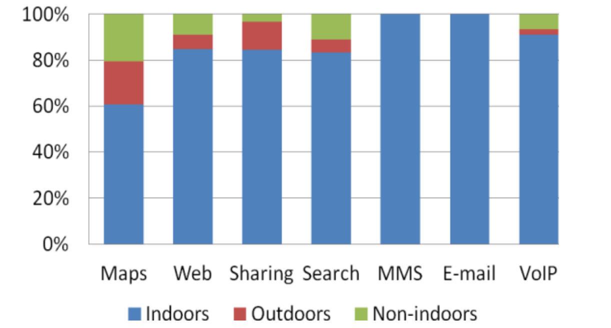


Figure 7. P009: Application Type Per Location Type

3.5 Correlating Connectivity and Location State for an Application Activity

After the connectivity as well as the location status analysis, we are correlating the network type used while using a specific application, with the location status of the user.

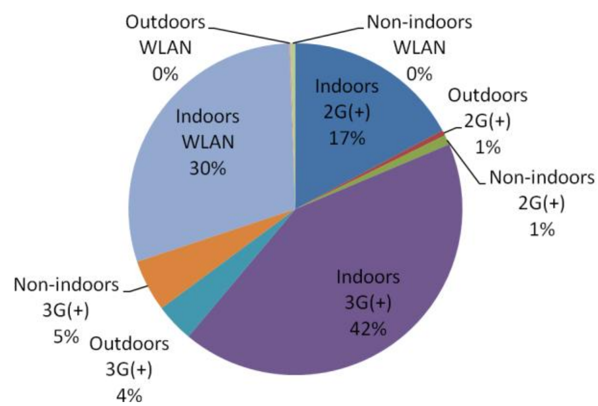


Figure 8. Connectivity Type vs Location Type (All Users)

Looking into the above chart, it is being clear the dominant case, when outdoors, is mainly the 3G(+) network and less the 2G(+). WLAN application activity in outdoors (or "non-indoors") location states is significantly low (almost zero when presented in the overall activity pie), meaning that users were extremely rarely connected to WLAN AP when out of indoor locations. When indoors, as one would expect, the WLAN percentages rise significantly, up to 30%.

Given the fact that we analyze mobile applications, we decided to focus on the detailed analysis of user application usage and connectivity of the *outdoors* and *non-indoors* location states, as presented in Figure 9.

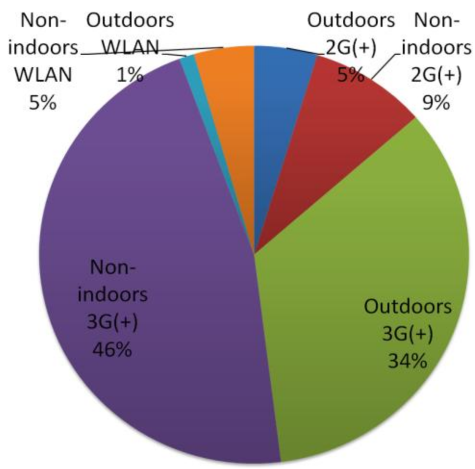


Figure 9. Application Type vs Location Type (All Users)

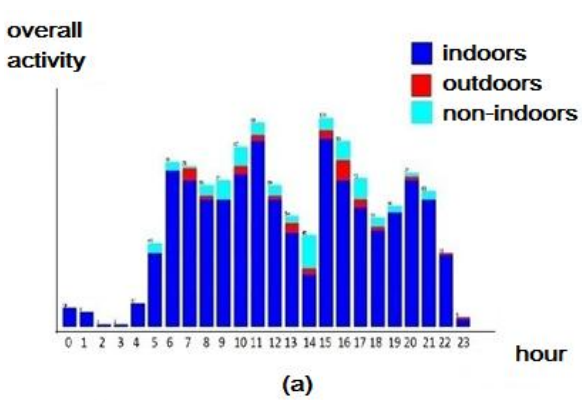
What we first conclude from Figure 9 is that users, when not in an indoor location, are well-connected (80% of the time) to the 3G network, which, as we know from literature, exhibits a higher performance than the 2.5G network.

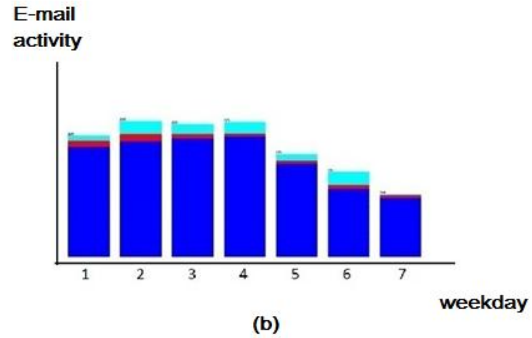
Moreover, the average WLAN application activity when not in an indoor location, is actually 6%, i.e., the outdoors state (1%) could be assigned to the case, where the uses is in a moving bus with WLAN connectivity, while non-indoors WLAN state (5%) could be assigned to the situation, where the user is outdoor, but close to the building or in a public space, where the WLAN is provided.

3.6 Temporal Activity

Finally, we analyzed the temporal aspect of the application activity. It seems that there are no differences per online application type and location status during one random day in the study. Users seem to be more active during work hours (8h-18h), with an exception during early afternoon, like 14am (Figure 10a).

As far as the days of the week are concerned, different types of applications show different temporal activity pattern. For example, *E-mail* - an application used by many people mainly for professional reasons - is characterized by a descending pattern while getting closer to the weekend (1-Mon, 7-Sun, Figure 10b), whereas, usage of *Maps* (Figure 10c) tend to increase during weekends, when, we hypothesize that people would change their usual mobility patterns (*e.g.*, going for a trip).





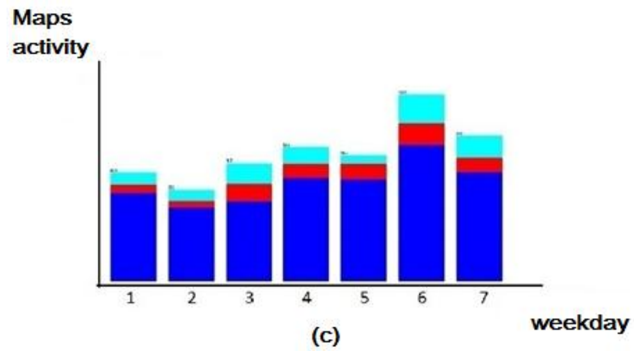


Figure 10. (a) Typical application usage per hour of the day, (b) *E-mail* and (c) *Maps* activity per weekday

4 DISCUSSION

In order to be able to label in an accurate way the different states of the mobile applications' activity and network connectivity, and the location state of the study participants, several assumptions have been made. The (lack of some) MDC data itself imposes some limitations on the model assumed for the data analysis. For instance, the absence of available information about the WLAN APs to which the user is connected, leads us to the need for assumptions, with the risk of results' inaccuracy.

Our choice for application filtering and clustering had to do with the goal of analysing the mobile user's patterns in the context of network demanding applications. We assume that the data requirements for different applications of the same context (*e.g.,* all VoIP apps observed) are approximately the same, thus their usage patterns can be analyzed as one. As mentioned earlier, however, the different applications' clusters are based on own research and distinctive, known traffic models, however, with diverse data requirements with one each other. The assumptions taken by us in the analysis may lead to inaccurate conclusions.

For the external validity of the above presented results, it should be noted that the correlation between the application activity patterns and the network connectivity, or the location states, was developed upon the fact that the analysis was carried out on how the participants *actually acted* during the study, given all the constrains on, *e.g.*, their data subscriptions, and possibly on the application's battery usage and their current battery status. The outcome might have been significantly different if additional methods had been emplyed during the study, aiming to capture what participants wished to do at a given moment, and not only what they did. This could be captured employing Experience Sampling Method [11] – presenting on the mobile phone screen a simple questionnaire to the user asking for his inputs for application usage needs and expectations, as well as experiences (after an application usage).

5 CONCLUSIONS

In this paper, we have presented an approach for deriving wireless access network connectivity and location patterns during mobile applications usage, and correlating them with basic location states of the mobile user, as well as the temporal aspects of it. After choosing certain clusters of mobile network-demanding applications, in this paper we have described a way several phone users interact with these applications of their phone, depending on the network they are connected to, and whether they are indoors or (possibly moving) in an outdoor environment.

The emerging requirement for all kind of mobile application activity is that one shall always attempt to ensure the Quality of Experience (QoE) level as expected by the mobile user. The results show, that, certain conditions application usage patterns may vary between applications as well as between users, possibly because of their different needs in different contexts. Nevertheless, the temporal and spatial aspect, as well the current connectivity options for the users seem to significantly impact their online application activity. For example, from the data one may roughly conclude that users are using applications on average 5% of the time when outdoors, while the detailed analysis revealed that for certain applications like Maps, this percentage raises up to more than 40 %.

The 70% of the mobile application activity is taking place under cell network (2/3G(+) type) and the rest under WLAN. WLAN activity is exhibited 99% of the time while indoors. From the temporal behaviour aspect, participants were active from 5am to 22pm and the activity intensity decreased from weekdays to weekend, apart from the specific application types like Maps, which activity increased over weekend, *i.e.*, possibly when a user was navigating outdoors in unknown places.

As future work, we would be interested in analyzing further the semantic places (*e.g.*, home, office) and correlating these places with the different application usage patterns (connectivity and temporal behaviour). Furthermore, the user's expected QoE could be linked to these patterns and prediction mechanisms could be developed to maximize this QoE either by choosing automatically the best network to be connected to, for the given mobile application, or by suggesting the set of applications to be used the user, based on the user's given wireless access network connectivity and predicted QoE for these applications.

6 ACKNOWLEDGMENTS

Support of MyGuardian (AAL-2011-4), WayFiS (AAL-2010-3-014) and TraiNutri (AAL-2009-2-129) projects is acknowledged.

7 REFERENCES

[1] H. Verkasalo, H. Hämmäinen (2007): A handset-based platform for measuring mobile service usage. Emeral Journal, 9(1), pp. 80 – 96.

[2] K. Wac (2009): Collaborative Sharing of Quality of Service-Information for Mobile Service Users, PhD thesis, SYSINF, University of Geneva, Geneva, Switzerland

[3] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R.Govindan, D. Estrin (2010): Diversity in Smartphone Usage, ACM MobiSys.

[4] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen (2012): The Mobile Data Challenge: Big Data for Mobile Computing Research,, in Proc. Mobile Data Challenge by Nokia Workshop, in conjunction with Int. Conf. on Pervasive Computing, Newcastle, UK.

[5] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica, Perez, J. Laurila (2010): Towards rich mobile phone datasets: Lausanne data collection campaign, ACM Intl. Conf. on Pervasive Services (ICPS), Jul. 2010.

[6] Nokia Support – "How to use Maps offline": http://europe.nokia.com/support/product-support/maps-support/how-to/how-to-use-maps-3.03/how-to-use-maps-offline, visited on 3/3/2012.

[7] Nokia Support, "Know your mobile", http://www.knowyourmobile.com/nokia/nokian95/nokian95internet/2533/setting_up_email_on_the_nokia_n95.html, visited on 29/3/2012.

[8] GSM for Dummies, "Network Architecture, GSM Network Architecture", http://gsmfordummies.com/architecture/arch.shtml#bts, visited on 02/3/2012.

[9] Google API, http://www.**google**.com/**glm**/**mmap**?mcc=xxx&mnc=xxx&towerid=xxx, visited on 28/5/2012.

[10] S. Ickin, K. Wac, M. Fiedler, L. Janowski, J.H. Hong, A.K. Dey (2012): Factors Influencing Quality of Experience of Commonly-Used Mobile Applications, IEEE COMMAG,, 50(4): 48-56, IEEE Press.

[11] J. M. Hektner, et al. (2006): Experience Sampling Method, Measuring the Quality of Everyday Life, Sage Pubs Inc.