

de Reuver, Mark; Bouwman, Harry

Conference Paper

Preferences in data usage and the relation to the use of mobile applications

25th European Regional Conference of the International Telecommunications Society (ITS): "Disruptive Innovation in the ICT Industries: Challenges for European Policy and Business", Brussels, Belgium, 22nd-25th June, 2014

Provided in Cooperation with:

International Telecommunications Society (ITS)

Suggested Citation: de Reuver, Mark; Bouwman, Harry (2014) : Preferences in data usage and the relation to the use of mobile applications, 25th European Regional Conference of the International Telecommunications Society (ITS): "Disruptive Innovation in the ICT Industries: Challenges for European Policy and Business", Brussels, Belgium, 22nd-25th June, 2014, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/101437>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Preferences in data usage and the relation to the use of mobile applications

Mark de Reuver - Delft University of Technology

Harry Bouwman – Delft University of Technology and IAMSR, Abo Akademi University

Abstract

While most user studies on mobile telecommunications focus on adoption of services, preferences in the use of data networks has hardly been studied. In this paper, we analyse data collected directly on smartphones to study preferences of users between cellular and WiFi networks. Moreover, we assess how the use of specific types of applications contributes to data consumption. In absolute terms, use of WiFi is higher than use of cellular networks. The spread among participants in use of cellular networks is very high, ranging from 0 to 100% of their total traffic. There are no significant differences between Apple and Android users. No effects were found of the size of the data plan on the amount of cellular data being consumed. Cellular network usage is especially driven by chat, social networking and browsing applications. High users of video applications do not significantly consume more bandwidth, which is at odds with conventional ideas on the capacity crunch. Log data on application usage explains data consumption better than self-reported usage levels. The results are relevant for telecom operators to steer the amount of data being consumed over their cellular and WiFi networks. However, data consumption levels do differ greatly across the population, and as a large proportion of data traffic cannot be explained by application usage levels, preferences for data usage cannot be very well explained.

Keywords: mobile services, mobile telecommunications, cellular networks, bandwidth, WiFi networks

1. Introduction

Most user studies on mobile telecommunications focus on the applications or services being used and seldom on “network behaviour”, i.e. who makes use of WiFi versus cellular data networks. In practice, however, network operators have to deal with load balancing on their networks without having a clear idea on what type of application is driving mobile data traffic. Some operators intend to stimulate usage of their 3G and 4G networks, while others are looking to reduce usage levels. Industry analysts often point out that only a small proportion of the mobile users consume most of the bandwidth available. Offloading to WiFi can be a solution as demonstrated in simulation studies (Dimatteo, Hui, Han, & Li, 2011).

For all these operator policies, insight is required into what drives data consumption patterns of users. The main research question in this paper is therefore: which applications drive mobile data usage on cellular versus WiFi networks?

Research on mobile data is predominantly based on questionnaire data, however with new technologies it is possible to measure actual behaviour (Verkasalo, 2007; Smura et al, 2009; Eagle & Pentland, 2006; Falaki et al, 2010; Boase & Ling, 2013; Kobayashi & Boase, 2012; Karikoski, 2013). In this paper we present the results of a two-year research project in which we experimented with combining both ways to data collection.

The paper contributes to existing research making use of log-data with a specific focus on data-usage.

Section 2 provides the method of the study, followed by results in Section 3. Section 4 concludes the paper and discusses the findings.

2. Method

We adopt an innovative quantitative mixed-method approach. To measure the use of mobile applications, we collect log data directly from the smartphone of 233 Dutch consumers. The log data contains detailed information regarding the time, duration and type of mobile application accessed by the user. Log data is to be preferred over self-reports of mobile application usage which tends to be biased due to unsystematic overestimation (Boase & Ling, 2013; Kobayashi & Boase, 2012).

2.1. Sample

The population comprises Dutch smartphone users of 16 years and above. As the measurement software only works with iPhone and Android smartphones, Symbian, Blackberry and Windows phone users are excluded from the study mainly because software development maintenance for the decreasing number of users of these platforms is no longer attractive. A user panel comprising 20,000 households was used to sample respondents. The user panel is representative for the Dutch population in terms of demographics. The panel is regularly renewed through active recruitment (i.e. no self-selection bias is involved) and panellists are typically not compensated for taking part in surveys.

From the panel, a random sample was drawn. Next, an initial questionnaire was sent to the persons in the sample inviting them to participate in the study. The initial questionnaire extensively explained how log data would be collected, stored and analysed in the study, as well as how privacy would be guaranteed (Bouwman, de Reuver, Heerschap, & Verkasalo, 2013). As the first round of recruiting did not lead to sufficient response, the procedure was repeated but only including the subset of respondents that were known to possess a smartphone. Finally, in a third recruiting round, panellists who participated in an earlier pre-test based study were also approached to participate.

After data cleaning for partial non-response, the three rounds of recruitment resulted in data from 1653 persons that filled in the initial questionnaire, out of which 519 (36%) were willing to participate in the study. A considerable number of respondents didn't own a smartphone (12%). A large part refused to participate (56%). Of the reasons for non-participation provided, the core reason was privacy (by 16% of the respondents). For 15% of the respondents the reasons were related to typical non-response reasons, such as holidays, sickness and travelling abroad. Technical reasons were mentioned by 2% of the respondents, and 3% indicated their employer would not allow them to download apps on their phones. Other reasons provided included low usage of the smartphone and no experience or cognitive capabilities to install applications on their smartphone.

Although 519 respondents initially indicated that they were willing to participate in the study, only a part of them downloaded and installed the app (369). Only 232 participated for the full four weeks of the study. Reasons to drop out during the study were related to technical problems, like battery drainage and reduced performance of the phone. Some respondents dropped out because they upgraded to a new version of their operating system. Furthermore there are the usual reasons like travelling abroad, sickness and so on. Sample characteristics are given in Table 1. On average, participants were 46 years old.

Table 1: Sample characteristics (N=232)

Gender	Male	51%
	Female	49%
Position in household	Primary wage owner	30%
	Caretaker	29%
	Both	32%
	Child	7%
	Other	3%
Education	HW	15%
	HB	41%
	HA	9%
	MB	23%
	MA	7%
	LB	4%
	LA	0%
Activity	Work:	71%
	Retired:	10%
	Caretaker:	4%
	Student:	7%
	Unemployed:	2%
	Incapable to work:	4%
	Social Benefit:	0%
	Other:	1%
Family size	1	20%
	2	40%
	3	15%
	4	18%
	>4	7%
Income	Below Modus	12%
	Modus	22%
	Above Modus	62%
	DNK, no answer	4%

2.2. Log data metrics

Our approach is similar to the few previous studies that utilize smartphones to automatically log user activities (Eagle & Pentland, 2006; Falaki, Mahajan, & Kandula, 2010; Raento, Oulasvirta, & Eagle, 2009; Verkasalo & Hämmäinen, 2007). For an overview see Karikoski (2012).

To carry out smartphone measurement, a number of software tools are available, for example LiveLab (Shepard, Rahmati, Tossell, Zhong, & Kortum, 2011) and Device Analyzer (deviceanalyzer.cl.cam.ac.uk) (see for an overview also Karikoski, 2012). The present study utilizes the commercially available smartphone measurement application from Arbitron Mobile. The measurement application runs on the background of the mobile phone, and transmits log files daily to the server. The application can be downloaded from the app store. Participants were given the opportunity to view a dashboard with their personal usage numbers during the period of the study. The software was pretested in 2011; identified technical problems with the software were solved afterwards.

The software logged each action of the user over a period of 28 days (30 October – 27 November 2012). Each time an application is launched, the software logs the application name, date and time, and duration in which it is displayed on the foreground of the device. The software classifies applications into specific types using automated content analysis. The researchers manually checked and verified the most frequently used applications and found no errors.

3. Results

3.1. Data exploration

Data traffic can take place via an operator network (i.e. cellular network, which can be 2G or 3G) or a WiFi network. Overall, consistent with earlier research executed in 2011, more megabytes are transmitted through WiFi networks than through cellular networks. In addition, more data is received than sent via a smartphone. Figure 1 indicates the spread of data traffic across the sample, like minimum, maximum and median. When the scale is transformed logarithmically, only few outliers remain. There are more outliers for cellular networks than for WiFi networks, which indicate that the spread is larger. The data shows that there is a considerable amount of individuals that hardly use cellular networks, or even not at all and only apply WiFi.

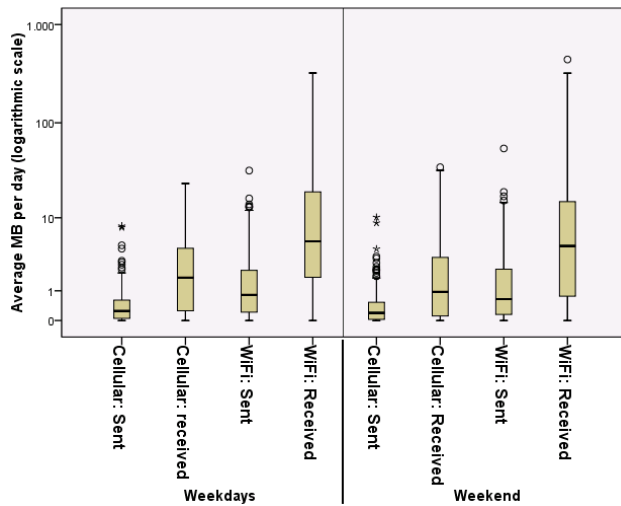


Figure 1: Data consumption per day

On average, about 30% of outgoing bytes and 27% of incoming bytes are transmitted via cellular networks. The boxplots in Figure 2 show again the large variation across the sample, with the range between 0% and 100%.

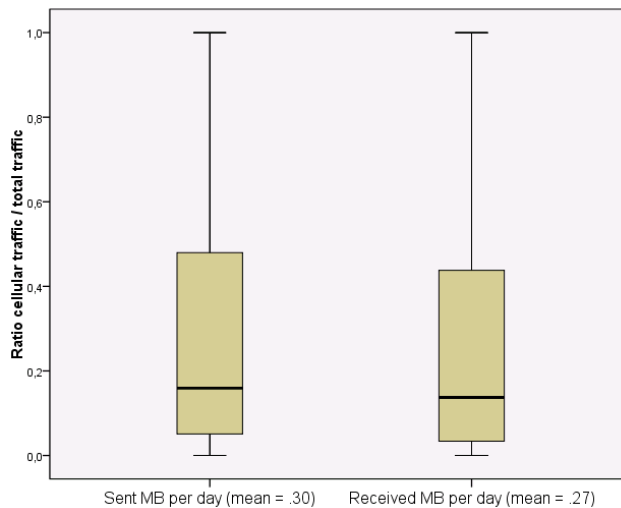


Figure 2: Proportion of data traffic (MB) via cellular networks (N=232).

There is strong correlation between the amount of MB's sent and received over cellular networks and over WiFi networks. In other words, heavy users of WiFi are also heavy users of cellular networks. See table 2.

Table 2 Correlation between amount of data traffic (MB) through Wifi and cellular networks (N=232).

	Cellular: sent	Cellular: received	WiFi: sent	WiFi: received
Cellular: sent	1			
Cellular: received	.63**	1		
WiFi: sent	.40**	.20**	1	
WiFi: received	.14*	.17**	.50**	1

** $p < .01$, * $p < .05$

The use of cellular networks on Android versus Apple is not significantly different, see Figure 3. However, the use of WiFi is significantly different for both MB sent ($t(231)=2.31$, $p = .022$) and received ($t(63)=4.05$, $p = .000$). Overall, it appears that Apple users use WiFi more than Android users.

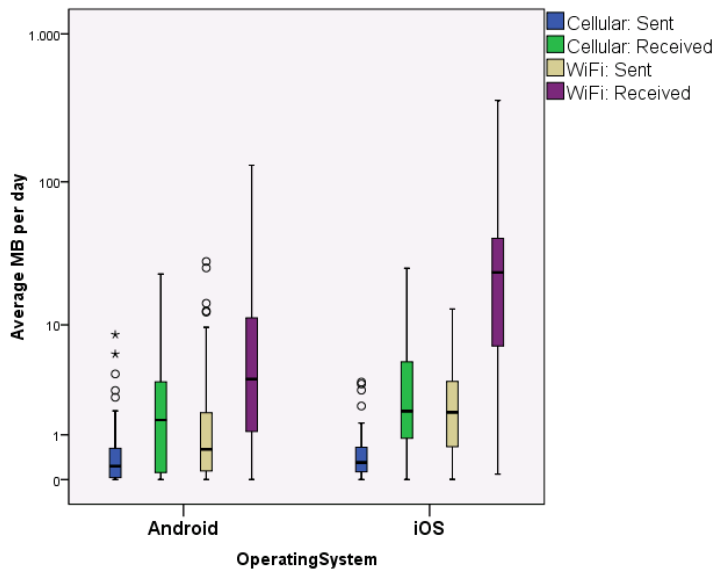


Figure 3 Data traffic per day, compared between operating systems

We also tested mean differences of data usage patterns between different demographic groups. However, gender, age, income levels, education levels and family size do not affect data traffic significantly in the sample.

Similarly, the presence of a data plan and the cap on the data plan does not correlate significantly with the amount of data used through cellular or WiFi networks.

3.2. Correlation between use of applications and data traffic

It can be expected that data traffic and use of applications are interrelated. Figure 4 shows the number of minutes spent on the four major types of applications per day, and the total amount of MB sent and received on both cellular and WiFi networks per day. It is striking that the trend line for browsing, social networking and chat applications are more of less similar. For gaming the trend line is not very strong, likely because games are often used offline without heavy data traffic. Overall, explained variance is low.

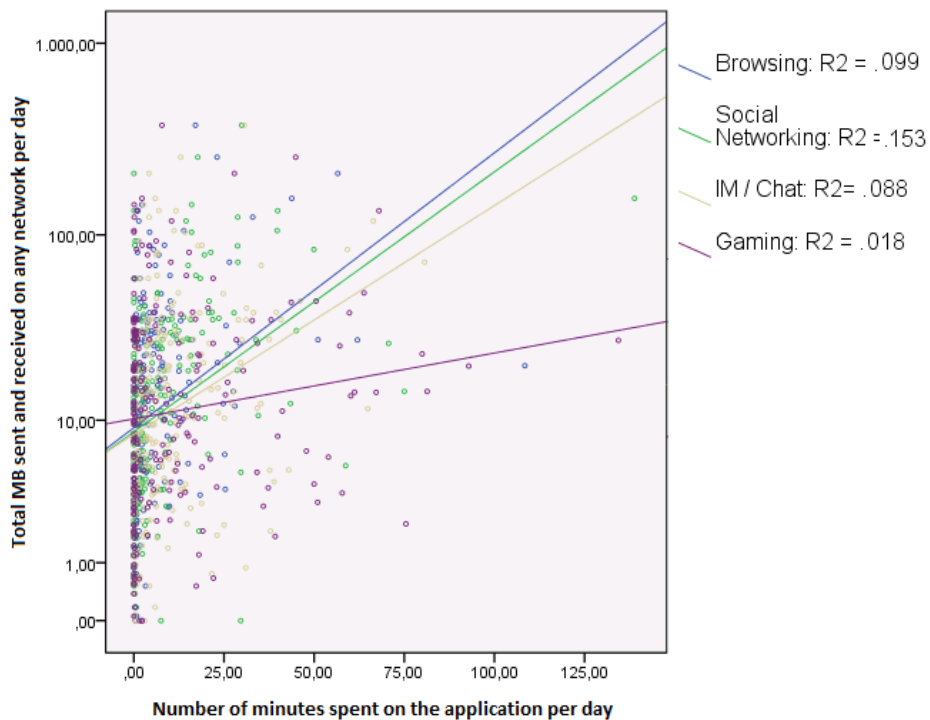


Figure 4 Scatterplot on application usage and data traffic

Table 3 shows a regression model that explains that the number of MB's by all four application types is significant ($F(4)=16.477$, $p = .000$; $R^2 = .227$). The effect size of social networking apps is the largest, followed by IM / Chat and browsing.

Table 3 Regression model explaining data traffic through application usage

	Beta	t	P	VIF
(Constant)		14.467	.000	
Social Networking Apps (avg minutes / day)	.270	4.134	.000	1.237
Browsing Apps (avg minutes / day)	.178	2.818	.005	1.160
Gaming Apps (avg minutes / day)	.105	1.778	.077	1.010
IM / Chat Apps (avg minutes / day)	.185	3.033	.003	1.080

When testing these patterns in structural equation model, we can see from Table 4 that cellular network traffic is driven by social networking, browsing and chat applications. Gaming and video have a negative impact. This is striking because the typical idea is that consumers using YouTube create an overload on the mobile network. Apparently, people that use gaming do not use their time with their smartphone on receiving or sending information, i.e. gaming mainly takes place offline. WiFi traffic is also driven by social networking and browsing applications, but not by instant messaging. Gaming and video have hardly an effect.

Table 4 Structural equation models testing impact of application usage on data consumption

	Cellular MB total sent & received	WiFi MB total sent & received
Video	-.15*	n.s.
Social networking	.15*	.17**
Browsing	.26***	.24***
Gaming	-.12*	n.s.
IM / Chat	.12*	n.s.
Productivity	n.s.	n.s.
Maps / Navigation	n.s.	n.s.
App store	n.s.	n.s.
Explained variance	.198	.116
Overall model fit	$\chi^2(3) = .070$, $p = .995$, $NFI = 1.00$, $CFI = 1.00$, $TLI = 1.22$, $RMSEA = .000$	$\chi^2(6) = 8.166$, $p = .226$, $NFI = .955$, $CFI = .985$, $TLI = .910$, $RMSEA = .039$

Finally, we test if we find different effects if we rely on self-reported levels of application usage rather than log data metrics. Table 5 shows that the total explained variance is lower for self-reported scales. Overall, similar patterns are found, i.e. social networking and browsing contribute to cellular usage, while gaming has a negative effect.

Table 5 Linear regression model testing impact of reported application usage on actual data consumption

	Cellular MB total sent & received	WiFi MB total sent & received
Mobile TV	n.s.	.14*
Social networking	.21**	n.s.
Browsing	.20**	n.s.
Gaming	-.17*	n.s.
IM / Chat	n.s.	n.s.
Productivity	n.s.	n.s.
Maps / Navigation	n.s.	n.s.
Explained variance	.141	.086

4. Discussion and conclusions

Data consumption patterns differ greatly within our sample of 233 smartphone users. Especially for cellular network usage, usage levels are highly diverse. Demographic variables, type of smartphone and size of data plan cannot explain cellular network consumption patterns at all. The use of applications can only explain a moderate part of the variance.

As such, the possibilities for telecom operators to understand and subsequently steer data consumption are limited. From the results, it appears that social networking, browsing and instant messaging drive the use of cellular networks. To reduce the load on cellular networks, telecom operators should thus either reduce the use of these applications or implement means to entice users to switch to WiFi networks for these applications. A striking finding is that video usage does not drive cellular consumption levels. Although it has been posited by industry analysts that the few users that consume video on the go consume the most of the cellular networks, this is thus not supported by our data. In other words, streaming video cannot be held responsible for high data

loads. If replicated, this finding would have important implications for telecom operators but also the net neutrality debate in telecommunications policy.

When comparing a model with log data as predictors and self-reported application usage as predictors, we found that log data performs better. This contributes to the argument that was posed earlier in literature to rely on log data rather than self-report survey data (De Reuver et al 2012; Boase & Ling, 2013).

Results provide insights for operators on the type of app usage that should be stimulated in order to increase data traffic. At the same time, the results provide insight into the trade-off that should be made between unlimited and limited data plans. In future research, the type of cellular network should be included as 4G behaviour may be radically different than 3G users.

A limitation of the present study is that the distinction between home, work/study and travel context is not taken into account. Earlier studies have identified such context variables from smartphone log data, but implementing the algorithms is challenging.

A next step could be to focus on the heavy users of cellular networks in a more qualitative study. As this study has attempted to cover a balanced sample of mobile users, such heavy users could be more prominent in a future study.

References

- Boase, J. & Ling, R., 2013. Measuring Mobile Phone Use: Self-Report Versus Log Data. *Journal of Computer-Mediated Communication*, doi: 10.1111/jcc4.12021
- De Reuver, M., Bouwman, H., Heerschap, N., & Verkasalo, H. (2012). Smartphone measurement: do people use mobile applications as they say they do? Paper presented to International Conference on Mobile Business, Delft, the Netherlands.
- Eagle, N. & Pentland, A. (2006). Reality Mining: Sensing Complex Social Systems. *Personal and Ubiquitous Computing*, 10 (4), pp.255-268.
- Falaki, H., Lymberopoulos, D., Mahajan, R., Kandula, S. & Estrin, D. (2010). A first look at traffic on smartphones. In: *Proceedings of the 2010 Internet Measurement Conference (IMC2010)*. Melbourne, Australia 1-3 November 2010.
- Kobayashi, T., & Boase, J. (2012). No such effect? The implications of measurement error in self-report measures of mobile communication use. *Communication Methods and Measures*, 6(2), 126-143.
- Karikoski, J. (2013). Empirical Analysis of mobile interpersonal communication service usage. Aalto University PhD thesis.
- Smura, T., Kivi, A. & Töyli, J., 2009. A framework for analysing the usage of mobile services. *Info*, 11 (4), pp.53-67
- Verkasalo, H., 2007. Handset-based measurement of smartphone service evolution in Finland. *Journal of Targeting, Measurement and Analysis for Marketing*, 16 (1), pp.7-25
- Dimatteo, S., Hui, P., Han, B., & Li, V. O. (2011, October). Cellular traffic offloading through WiFi networks. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on* (pp. 192-201). IEEE.