

## METHODOLOGY FOR ANONYMOUS MEASUREMENT OF AGGREGATED SMARTPHONE DATA TRAFFIC AND CHARACTERISTICS OF SMARTPHONE USERS

Siniša Husnjak, Dragan Peraković, Marko Periša  
Faculty of Transport and Traffic Sciences, University of Zagreb  
shusnjak@fpz.hr, dperakovic@fpz.hr, mperisa@fpz.hr

**Abstract:** *By using smartphones, users generate data traffic. Users' characteristics are an important factor in generating smartphone data traffic. This research presents a methodology for measuring smartphone data traffic and collecting data about individual characteristics of users of those smartphones. The measurement of generated data traffic is achieved by using an integrated smartphone application. The individual characteristics of smartphone users are determined on the basis of answers collected through the online questionnaire. The results of the developed methodology enable anonymous collection of aggregated smartphone data traffic and the characteristics of the users of those smartphones.*

**Key words:** *smartphone, data traffic, anonymization, measurement, methodology*

### 1. Introduction

Smartphones are connected with diverse broadband networks and provide their users a constantly increasing scope of services and applications and lead to explosive increase in the generation of data traffic. Since the possibility of data transfer of smartphones is becoming stronger and stronger, the attitude of telecommunication operators is changing as well. Today, for the first time the quality of data transfer service and pricing of the mobile data traffic are becoming more important than the voice transfer.

Increased use of smartphones and applications, which require access to network at any moment, has changed the user behaviour. The mobile network operators do not have the complete insight into the behaviour of the users and the generation of data traffic during the user access to Wi-Fi network. The mobile network operators are aware that the use of Wi-Fi networks is rooted in the user behaviour and usage, but the fragmentation of Wi-Fi networks has made precise measurement of such usage difficult. Connecting of smartphones to access Wi-Fi points and the data traffic generated in this way represents a part of generated data traffic that cannot be measured by the mobile network operator.

Smartphones offer the possibility of collecting information on the generation of data traffic by using applicative solution. Data can be used to obtain information about the habits and behaviour of the users. Literature states different methods of measuring the generated data traffic of smartphones. Every method has its advantages and drawbacks.

In case of using the application of third parties for measuring data traffic the drawback is the price of developing the application, the accuracy of measurement, the need for installation on the smartphone, privacy of data and the possibility of measuring data traffic only after the application has been installed on the device. This indicates a very complex, time-consuming and expensive methodology with limited number of participants regarding the method of implementation. A very big drawback of individual methods of measuring data traffic is also the impossibility of measuring data traffic generated through the access to Wi-Fi networks, which occurs often in case of measuring done by telecommunication operators in their own telecommunication network.

Regardless of the used method of measuring the smartphone generated data traffic, the users' characteristics are usually examined by using subjective methods – survey questionnaire or interview. In the mentioned case, it is of utmost importance to insure anonymity of the users and privacy of the collected data, and as such this represents an imperative in every methodology of collecting and analysing data. Further in the text, the existing studies of the respective area will be presented, as well as the elements and characteristics of the newly developed methodology.

## **2. Previous research**

A substantial number of authors have analysed the use of smartphones within various contexts, but a minor number of studies has dealt with the description of various methodologies of measuring the smartphone data traffic, generated by the access to mobile and Wi-Fi networks.

Some of the authors present methods of measuring the smartphone generated data traffic, without the description of the user characteristics. Research [1] shows various approaches to measuring mobile data traffic, and the objective and subjective methods with their characteristics have been defined. Research [2] represents a detailed overview of research methods, their advantages and drawbacks as well as the measuring points within the telecommunication network. In research [3], by using the developed Android application, the data have been collected and analysed in order to understand the characteristics of smartphone data traffic and study the relation between the context of usage of the smartphones.

Some of the authors analyse, apart from the mobile network, also the Wi-Fi network. Thus, research [4], by using the developed application, presents the trends in generating smartphone data traffic and related applications by the access to mobile and Wi-Fi networks, whereas research [5] with the developed application presents the weekly and daily distribution of smartphone data traffic by accessing mobile and Wi-Fi networks.

Some authors include also next to the data on generation of smartphone mobile data traffic also the characteristics of the users. Falaki in [6] presents a detailed analysis of the forms and amounts of the generated mobile data traffic generated by smartphone by using various information and communication services and applications. Ghose in [7] on a large number of respondents gives a presentation on the users' behaviour, based on

the generation of mobile data traffic, and He in research [8] analyses the forms of generating mobile data traffic in various smartphones for designing of necessary performances of the mobile network. Gerpott et al. in research [9] present the relation between the monthly generated mobile data traffic, personal characteristics of the users and four different types of smartphones (according to OS). This research uses the data on monthly generated mobile data traffic obtained by the mobile network operator. The mentioned studies do not include the generation of Wi-Fi data traffic.

In studies [10] and [11] the measurements on mobile phones by developed application are supplemented by the questionnaires filled in by the users who participated in the research. Soikkeli in [12] uses the measurement on mobile devices themselves for collecting information related to the period of using mobile phones. The data are collected by a group of test users who had the pre-installed software / application for the control of traffic. The data have been used to describe the context of using mobile devices, where they help in understanding the behaviour of the users. Kaiser in [3] compares the data traffic generated by the mobile and Wi-Fi networks, and the data are collected by using the application in order to be able to understand the characteristics of the participants in different contexts (network, location, time, etc.)

The drawback of previous studies can be seen in the fact that some studies indicate the generation of mobile and not Wi-Fi data traffic. Further, no one research uses the integrated application of smartphone Android OS. Additionally, few studies include the users' characteristics with the generation of data traffic and the procedure of anonymization is not particularly emphasised. As conclusion, further in the text a methodology will be presented, which takes into consideration several significant both elements and data while collecting the smartphone data traffic and the users' characteristics.

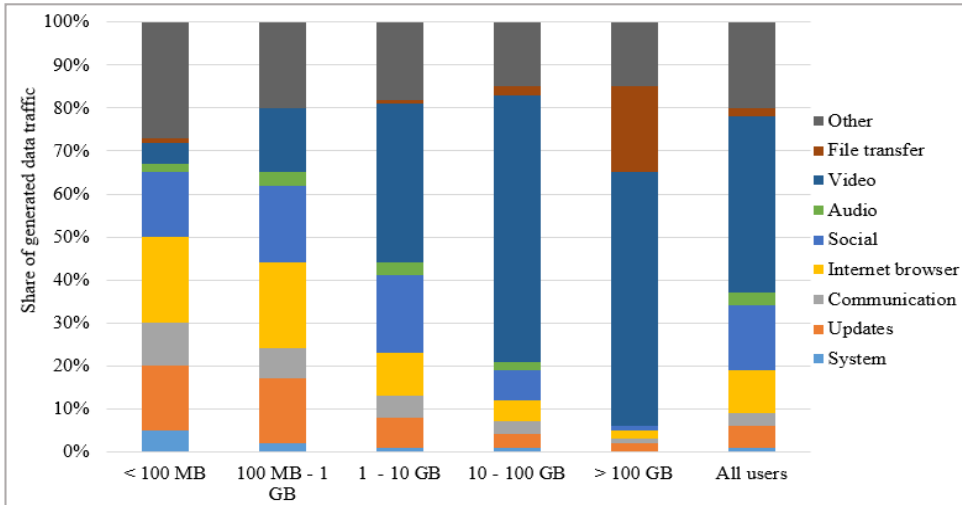
### **3. Generating and measuring smartphone data traffic**

Generating of smartphone data traffic includes:

- mobile data traffic – amount of data traffic generated by the smartphone using access exclusively via mobile networks;
- offload data traffic – amount of smartphone data traffic transferred from the mobile to Wi-Fi network i.e. generated by smartphone using access exclusively via Wi-Fi networks;
- data traffic – summarized amount of data traffic generated by smartphone using access to mobile and Wi-Fi networks.

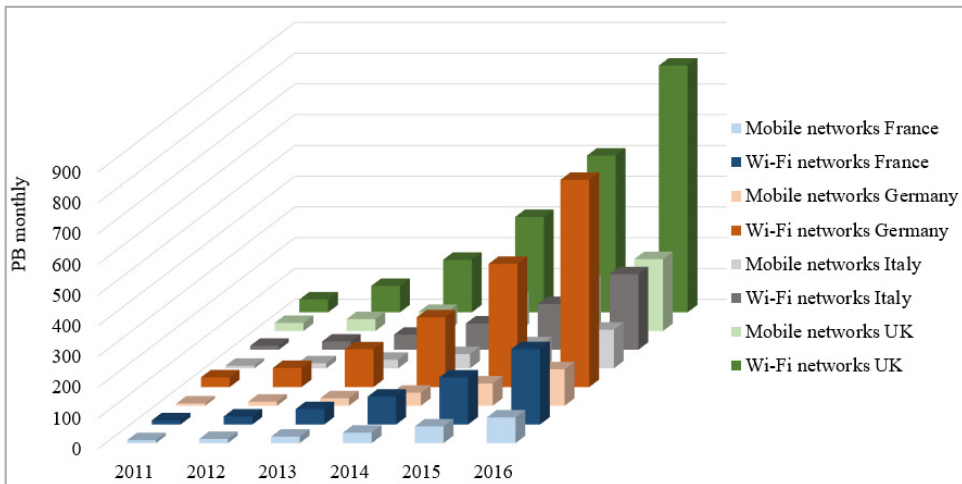
#### *3.1 Smartphone data traffic*

An increasing number of smartphone users is the main actuator of the increase of mobile data traffic. Because of the ever greater use of smartphones, the mentioned devices will generate three quarters of the total mobile data traffic by the year 2019 [13]. The generation of mobile data traffic continues to grow. This growth has been partly prompted by the growth of the popularity of video recordings, since the users rely on their mobile devices in using various services, as shown in Graph 1. The authors in [14] indicate that the reason for the growth of mobile data traffic is the availability of advanced smartphones which facilitate the usage of mobile services and attractive tariff plans.



Graph 1. The share of smartphone generated data traffic according to the types of used services and user categories, [14]

Smartphones are essentially meant for connection to the mobile network but their technological development has made it possible to use the Wi-Fi network as well. Wi-Fi technology has strongly formed the position of the most used wireless technology in the terms of the amount of generated data traffic that is transferred by the communication networks [15]. Although the generation of data traffic by using mobile networks and application of smartphones is on the rise, the growth of generated data traffic using Wi-Fi networks has exceeded it [16], which can be seen in Graph 2, for different countries of Europe.



Graph 2. Generated data traffic using mobile networks and the amount of data traffic offload from mobile to Wi-Fi networks, [17]

The monthly generated data traffic using Wi-Fi networks is much greater than the data traffic via mobile networks on 3G and 4G smartphones, and it is also growing much faster [18], [19].

### 3.2 Measuring the generated data traffic

Methods and elements of telecommunication networks that allow measuring of smartphone generated data traffic have been described in detail in researches [1], [2] and [20]. Research [20] explains in detail the advantages of measuring data traffic on the device itself (Handset-based measurement). The authors indicate that measurement on the device itself shows a number of advantages such as very good accuracy of measurement, quantitative and qualitative types of data that are collected, a wide variety of questions, etc. It should be emphasised that measuring of generated data traffic on the device itself enables the display of data traffic generated by the access to mobile and Wi-Fi networks. This then realizes the actual results that cannot be realised by other methods of measurement, regarding their method of measuring.

Integrated (original) application for monitoring data traffic is an integrated characteristic which allows the users to follow their generated data traffic according to single applications, whereas the applications of the control of the generated data traffic of the third party are available as applications that can be subsequently downloaded directly on the device, with the same functionalities.

## 4. Developed methodology for anonymous data collection

The developed methodology for collection of data on smartphone generated data traffic and the characteristics of users of these smartphones in this research include three main phases shown in Figure 1 and explained in further chapters:

1. measuring and collecting data on generated data traffic of Android OS smartphones using mobile and Wi-Fi networks;
2. collecting information on the characteristics of smartphone users;
3. anonymization and aggregation of the collected data.

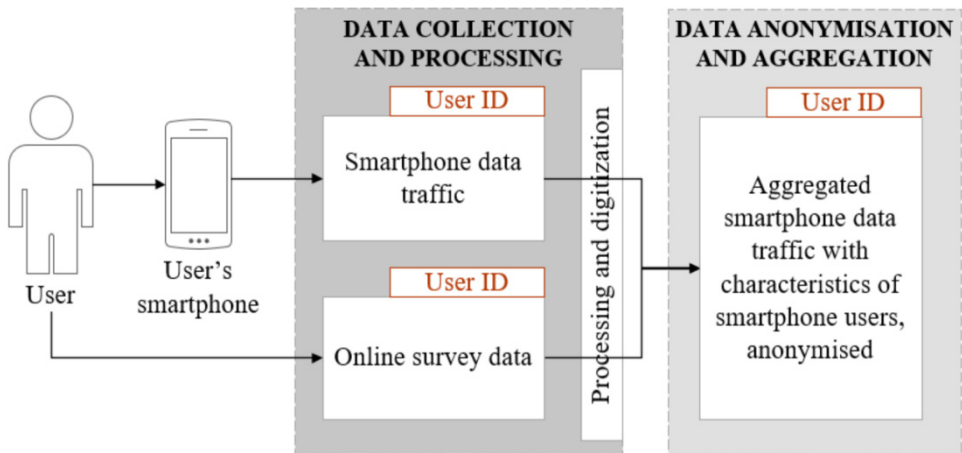


Figure 1. Participants and flowchart of the developed methodology

#### *4.1 Data collection and processing*

As can be seen in Figure 1, the process of collecting and processing data on the smartphone generated data traffic based on Android OS and the characteristics of users of these smartphones needs to be implemented based on the set of data from two sources:

1. filled in Form with data on generated data traffic by the use of mobile and Wi-Fi networks of users' smartphones, and
2. information on relevant characteristics of the users of these smartphones collected by online survey questionnaire.

Collecting data on the generated data traffic of the smartphone users has to be done by the method of interviewing the users. Based on the measurement through integrated application of Android OS smartphone, the historical data on the generated data traffic are collected according to individual applications. The functionalities of the integrated application include the possibilities of measuring the generated data traffic by using mobile and Wi-Fi networks. The application offers the possibility of selecting the observed time period of measurements, which depends on the type of research and the author. The measured data by integrated application are recorded into the Form that is handed to every research participant.

Furthermore, by online survey questionnaire the data on relevant characteristics of the smartphone user are collected, and they represent potential significant research variables. The questions of the survey questionnaire can include all the areas related to the interests and the objective of research; e.g. demographic characteristics of the user, type of subscription – prepaid/postpaid, used tariff plan, type and model of smartphone, possibilities of accessing Wi-Fi networks, etc. Online survey questionnaire is filled in on one's own smartphone via link available on the Form. The link is enabled based on QR or bar code, and the link is accessed by scanning the mentioned codes.

#### *4.2 Data anonymization and aggregation*

Individual smartphone user who participates in the research gets a unique identification key on the written Form. This key connects the collected data on smartphone data traffic and online survey questionnaire of the users by filling in the value of the obtained unique identification key as the initial field and it is the obligatory data of the online survey questionnaire.

The obtained unique identifier has a double role:

1. Anonymization of the collected data of the research respondents so that the collected data based on the used identifier cannot in any way be connected to the identity of the research participant;
2. Aggregation of the collected data of research participants so that the identifier used in the Form and filled in survey questionnaire connects the data of one user into a unique set of data on the generated data traffic and user's characteristics.

### **5. Discussion**

This research presents the new methodologies for measuring the generated smartphone data traffic based on Android OS and the characteristics of the users of these smartphones with the following most important characteristics:

- Usage of integrated application which is installed in every smartphone based on Android OS;
- Measuring of summarized amount of generated smartphone data traffic by access to mobile and Wi-Fi networks;
- Display of the characteristics of users of analysed smartphones based on their responses to the questions of the online survey questionnaire;
- Anonymity of the users of smartphones realised on the basis of unique identification key.

One of the important conclusions of this research is the recognition of the facts that the mobile network operators do not have full data and the visibility regarding generation of data traffic on smartphones by their users.

Compared to other methodologies of measuring data traffic the fact should be emphasised that the usage of integrated application allows measuring of generated smartphone data traffic which includes access to mobile and Wi-Fi networks. Since 86.1% of smartphone users use Android OS [21], to the majority of users the integrated application for measuring data traffic is available. The use of unique identification key allows a simple procedure of anonymization of the users and collected data, and the questions in the survey questionnaire insure modularity of research since those users' characteristics are examined which the researcher considers necessary and significant for the research.

The restrictions of the presented research refer to the fact that the methodology, regarding the characteristics of integrated application for measuring smartphone data traffic, is applicable exclusively to smartphones based on Android OS. The restriction of the methodology is also time-consuming process of collecting and processing data. This can be optimised by using the application that would have the possibility of extracting and saving of data from the integrated application on the generation of smartphone data traffic.

The practical application of the results obtained in this work are directed first of all to the mobile network operators, mobile virtual network operators, mobile network service providers, Wi-Fi network operators and Wi-Fi network service providers with the aim of providing information on the potential and possibilities of using the developed methodology in order to realise a more detailed insight into the behaviour of users in generating data traffic using smartphones.

## **6. Conclusion**

Analysing more deeply the trends and making decisions by using the insight into the data on the users, the operators have the opportunity of using more efficiently the total demand for data traffic by using mobile and Wi-Fi networks. This can be realised by finding innovative methods for measuring the realised data traffic and the characteristics of the smartphone users.

This research has made an important step towards the efficient collection and anonymization of aggregated data on the amounts of offloaded smartphone data traffic with survey data on the characteristics of the users of these smartphones.

The developed methodology has shown that the data on the quantities of generated smartphone data traffic can be aggregated with the data on the users'

characteristics based on the unique common identifier, thus achieving the users' anonymity.

The data aggregated in this way can be used as the base for the development of the model which is used to define the patterns of users' behaviour, providing knowledge for the needs of strategic action and optimization of the business of telecommunication operators in terms of personalisation of services and tariff plans, segmentation and targeting the users as well as planning of the telecommunication network. The mentioned level of insight in adopting information and communication services on smartphones and understanding how they are used on different access networks will be critical for the development of the future pricing, use of information and communication services and partner strategies between different participants of the dynamic telecommunication market.

The mobile network operators, based on the available methodology can realise the insight in the users' behaviour and quantities of the data traffic that are generated by using mobile networks, which are offloaded from mobile networks to Wi-Fi networks. If the operators have the possibility of understanding the differences in using various access networks by their users, they can use the mentioned insights for the construction of the user targeted proposals. For instance, for the construction of dedicated tariff plans for popular applications, such as e.g. unlimited e-mail plan for the roaming services or data package with included free access to popular social networks.

The operators of telecommunication networks should direct their efforts to proactive influence and design of users' decisions with price modification and tariff plans, user engagement and network planning strategy, and not imposing arbitrarily the rules that can go contrary to deeply rooted behaviour of users. The capability of accurately forecasting the quantity of generated and offloaded data traffic is crucial for the management of the growth of data transfer. The problem is that not all devices are equal, and data traffic varies significantly for every network on the basis of available applications in this network. A systemic approach is necessary for the analysis of the influence of data services regarding the very behaviour of the user in order to realize their better forecasting and greater added value of the service provider.

This research opens up new areas on the possibilities of collecting data on the generated data traffic and the characteristics of smartphone users, in order to optimise the business of operators and providers of services regarding access to mobile and Wi-Fi networks.

## Literature

- [1] T. J. Gerpott and S. Thomas, "Empirical research on mobile Internet usage: A meta-analysis of the literature", *Telecommunications Policy*, vol. 38, no. 3, pp. 291-310, 2014.
- [2] A. Kivi, "Measuring Mobile User Behavior and Service Usage: Methods, Measurement Points, and Future Outlook", *International Journal of Mobile Communications (IJMC)*, vol. 7, no. 4, pp. 415-435, 2009.
- [3] S. Kaiser, "Smartphone traffic characteristics and context dependencies", Master of Science Thesis, University of Saskatchewan, Saskatchewan, Canada, 2012.
- [4] Informa telecoms and media, "Understanding today's smartphone user: Demystifying data usage trends on cellular & Wi-Fi networks", Report,



- Vancouver, British Columbia, 2012.
- [5] J. Lee and P. Seeling, "An Overview of Mobile Device Network Traffic and Network Interface Usage Patterns", *Proceedings of the IEEE International Conference on Electro/Information Technology (EIT)*, pp. 0-4, 2013.
- [6] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, "A first look at traffic on smartphones", *Proceedings of 10<sup>th</sup> annual conference on Internet measurement and analysis- IMC*, vol. 10, pp. 281, 2010.
- [7] A. Ghose and S. P. Han, "An Empirical Analysis of User Content Generation and Usage Behavior on the Mobile Internet", *Management Science Journal*, vol. 57, no. 9, pp. 1671-1691, 2011.
- [8] X. He, P. P. C. Lee, L. Pan, C. He, and J. C. S. Lui, "A panoramic view of 3G data/control-plane traffic: Mobile device perspective", *Lecture Notes of Computer Science (LNCS)*, vol. 7289, no. 1, pp. 318-330, 2012.
- [9] T. Gerpott, S. Thomas, and M. Weichert, "Characteristics and mobile Internet use intensity of consumers with different types of advanced handsets: An exploratory empirical study of iPhone, Android and other web-enabled mobile users in Germany", *Telecommunications Policy*, vol. 37, pp. 357-371, 2013.
- [10] T. Smura, A. Kivi, and J. Töyli, "A framework for analysing the usage of mobile services", *Info*, vol. 11, no. 4, pp. 53-67, 2009.
- [11] H. T. Verkasalo, *Handset-based Analysis of Mobile Service Usage*. Helsinki, Finland: Espoo, 2009.
- [12] T. Soikkeli, "The Effect of Context on Smartphone Usage Sessions", Master of Science Thesis, Aalto University School of Science, Aalto, 2011.
- [13] Cisco, "Cisco Visual Networking Index: Forecast and Methodology 2014-2019", Report, San Jose, California, 2015.
- [14] Ericsson, "Ericsson Mobility Report (June 2015)", Stockholm, Sweden, 2015.
- [15] Wireless Broadband Alliance, "Global Developments in Public Wi-Fi", Report, London, UK, 2011.
- [16] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020", Report, San Francisco, USA, 2016.
- [17] J. S. Marcus and J. Burns, *Study on Impact of traffic off-loading and related technological trends on the demand for wireless broadband spectrum*. Luxembourg, 2013.
- [18] Informa Telecoms & Media, "Smartphone use transforming with the rise of 4G and Wi-Fi", Report, Vancouver, British Columbia, 2014.
- [19] S. Husnjak, I. Forenbacher, D. Peraković, and M. Periša, "Preferences of Smartphone Users in Mobile to Wi-Fi Data Traffic Offload", *The Thirty-Fourth Symposium on Novel Technologies in Postal and Telecommunication Traffic*, vol. 34, pp. 295-305, 2016.
- [20] S. Husnjak, D. Peraković, and I. Cvitić, "Smartphone Data Traffic Measurement", *24th International Symposium on Electronics in Transport- ISEP 2016*, vol. 24, pp. 1-10, 2016.
- [21] Gartner. (2017). Worldwide Smartphone Sales to End Users by Operating System in 1Q17. *Gartner Says Worldwide Sales of Smartphones Grew 9 Percent in First Quarter of 2017* [Online]. Available at: <http://www.gartner.com/newsroom/id/3725117>.

**Sadržaj:** *Korišćenjem pametnih telefona korisnici generišu saobraćaj podataka. Pri tome, specifične karakteristike (navike) korisnika predstavljaju bitan faktor za generisanje saobraćaja. Ovaj rad prikazuje metodologiju za merenje saobraćaja i prikupljanje podataka o ponašanju individualnih korisnika pametnih telefona. Merenje saobraćaja vrši se putem integrisane aplikacije u pametnom telefonu. Specifične karakteristike individualnih korisnika prikupljaju se putem on-line upitnika. Razvijena metodologija omogućuje anonimno prikupljanje podataka o generisanom agregatnom saobraćaju podataka i navikama korisnika pametnih telefona.*

**Ključne reči:** *pametni telefon, saobraćaj podataka, anonimizacija, merenje, metodologija*

## **METODOLOGIJA ZA ANONIMNO MERENJE AGREGATNOG SAOBRAĆAJA PODATAKA I SPECIFIČNIH NAVIKA KORISNIKA PAMETNIH TELEFONA**

Siniša Husnjak, Dragan Peraković, Marko Periša