WILEY | Hindawi

Research Article

Data Traffic Offload from Mobile to Wi-Fi Networks: Behavioural Patterns of Smartphone Users

Siniša Husnjak (ID)**, Dragan Peraković** (ID)**, and Ivan Forenbacher** (ID)

Faculty of Transport and Traffic Sciences, University of Zagreb, Vukelićeva 4, 10000 Zagreb, Croatia

Correspondence should be addressed to Siniša Husnjak; sinisa.husnjak@fpz.hr

This paper presents a model for defining the behavioural patterns of smartphone users when offloading data from mobile to Wi-Fi networks. The model was generated through analysis of individual characteristics of 298 smartphone users, based on data collected via online survey as well as the amount of data offloaded from mobile to Wi-Fi networks as measured by an application integrated into the smartphone. Users were segmented into categories based on data volume offloaded from mobile to Wi-Fi networks, and numerous user characteristics were explored to develop a model capable of predicting the probability that a user with given characteristics will fall into a given category of data offloading. This model may prove useful for analysing smartphone user behaviour when offloading data.

1. Introduction

The number of smartphone subscription connections at the global level has reached 3 billion, and in the last five years, data traffic has increased more than 40-fold [1–4]. According to a Cisco report [5], smartphone-based data traffic is expected to exceed 80% of total data traffic generated on mobile networks by 2020. The increased use of smartphones, adaptation of price plans, and user demands for entertainment services and video content are accelerating the increase in data traffic offloading, and this growth is expected to continue [6, 7]. Data traffic offloading refers to traffic via complementary access network technology (e.g., Wi-Fi, WiMAX) instead of traffic via mobile networks [2, 8, 9].

This paper considers only offloading involving Wi-Fi networks [10], which account for a substantial proportion of data traffic offloading. More than 180 million Wi-Fi access points are spread around the world [11]. Wi-Fi technology is popular because it does not require licenced frequency spectrum and provides users with mobility to move within the local area coverage [12]. The quantity of data traffic that is currently being offloaded to Wi-Fi networks, primarily because of Wi-Fi access points in households, already exceeds the total amount of data traffic in mobile networks, and this proportion is expected to keep growing [6, 13]. A Cisco report

[5] concludes that the amount of smartphone-based data traffic offloading to Wi-Fi networks will be 56% of total data traffic by the year 2020, and studies have claimed that 65% [14] or more than 70% [15] of smartphone data traffic is being offloaded to Wi-Fi networks. One study of smartphone use found that the ratio of data traffic generated using Wi-Fi and mobile networks was greater than 4 : 1 [16]. At the global level, data traffic on mobile devices occurs via Wi-Fi networks and mobile networks in the ratio 80%/20% [17]. Smartphones and phablet devices will continue to dominate the generation of mobile data traffic [5].

Mobile network operators are aware that the use of Wi-Fi networks is a key element in smartphone users' behaviour. Measuring such use is challenging in part because of the fragmentation of Wi-Fi networks. As a result, mobile network operators cannot gain a detailed view of smartphone connection with Wi-Fi access points or of data traffic offloading.

Using applications integrated into the smartphone itself offers the possibility of measuring data traffic offloaded onto Wi-Fi networks, which would provide, for the first time, the opportunity for mobile network operators to measure this aspect of user behaviour. This could lead to insights that could guide the design of strategic actions, optimisation of operations (e.g., personalisation of services and price

plans), market segmentation, user targeting, and planning of telecommunication networks based on potential implementation of complementary Wi-Fi access networks.

2. Overview of Previous Studies

Many authors have analysed the use of mobile phones within various contexts, but few have examined the reasons for the substantial differences among users in the amount of smartphone data traffic offloaded [20–22].

Several studies have considered the generation of data traffic exclusively involving mobile networks, and this literature has established connections among user behaviour, smartphones, and particular information-communication services. Alwahaishi and Snášel [23] analysed user behaviour, characteristics, and intentions regarding the use of mobile Internet. Falaki et al. [24] analysed in detail patterns and volumes of mobile data traffic generated by smartphones during the use of various information-communication services and applications, such as Internet searching, instant messaging, and use of navigational maps. Baghel et al. [25] analysed the volume of data traffic generated by various smartphone apps. Ghose and Han [26] surveyed a large sample of users to derive a detailed view of their behaviour during the use of information-communication services and applications. Similarly, He et al. [27] analysed patterns of data traffic generated with various smartphones, in this case with the goal of achieving desired performance levels in mobile networks. Yang et al. [28] established a link among the volume of data traffic generated, smartphone-based apps, and user mobility. Binde and Fuksa [29] examined the development of mobile communication networks and information-communication services as factors affecting mobile Internet use in Latvia.

Other studies demonstrate possibilities for measuring data traffic generation on mobile phones or smartphones. One study [20] illustrated various approaches to such measurement, each with its advantages and drawbacks. Another study [30] laid out in detail research methods for such measurement and described their applicability to given research objectives. That work explored the concept of measurement points inside telecommunication networks as well as the possibility of predicting smartphone user behaviour and use of information-communication services.

Soikkeli [31] analysed periods of mobile phone use based on measurements collected on the mobile devices themselves. The data were collected by traffic-monitoring software previously installed on the devices of test users, and they captured the context in which the mobile phones were used, contributing to the understanding of user behaviour. Another study [32] also collected data on the use of mobile services using an application on the mobile device itself, which were complemented with data collected from a user survey. An app-based study analysed the distribution of smartphone-based data traffic on mobile and Wi-Fi networks across different times of day and on different days of the week [33], providing insights into working habits and behavioural patterns of smartphone users.

Few studies have examined how smartphone users access Wi-Fi networks, how they would want to use them in the future, and what makes them prefer to connect their device to Wi-Fi rather than to the mobile network [1]. One study [34] used an application for iOS and Android operating systems to analyse how the users of smartphones access networks (3G, 4G, and Wi-Fi access). Their results suggest that users select the network based on their access to newer communication technologies, their possibilities for data transfer speed, and economic possibilities. Smartphone user behaviour changes depending on the access network selected [16], and mobile network subscribers often choose to access Wi-Fi rather than a 3G or 4G mobile network [15]. This has given rise to the offloading "blackhole": when mobile subscribers switch from their mobile network to Wi-Fi, most operators can no longer track their activities, their data traffic amount, the quality of their user experience, or the security of their Internet connection [35]. This blackhole represents a large proportion of mobile subscriber behaviour: one study [34] indicates that users increase their data traffic when they move from a mobile to Wi-Fi network, and more than 40% of users connect to more than one Wi-Fi access point in a single day. This data traffic offloading creates several challenges for telecommunication operators, for which some authors have already proposed some solutions [8]. Bakhit et al. [36] concluded that offloading to Wi-Fi networks is an optimal offload option since it takes advantage of Wi-Fi availability and throughput capacity. Indeed, the data traffic offload from mobile to Wi-Fi networks has been stimulated by "free" availability of Wi-Fi access at homes, simplicity of automatic selection of the Wi-Fi network by smartphones, quality of experience (QoE) on mobile networks, and price structure of data services of mobile networks [16]. Hetting [37] states that the majority of users consider Internet access via Wi-Fi networks as less expensive, simpler, faster, and more reliable than access via mobile networks.

One approach to researching the blackhole of data traffic offloading is to use applications installed on the mobile device itself to measure the traffic (e.g., [38]). Taylor et al. [1] analysed user behaviour, smartphone use, and interest in accessing Internet via Wi-Fi networks. Kaisar [39] compared data traffic generated via mobile or Wi-Fi networks with an application that collected data on user characteristics such as network, location, and time. A potentially more informative approach is to supplement measurements taken on users' mobile devices themselves with data from surveys of the same users [32, 40].

Previous studies have documented the diversity of Internet access and of data traffic generation and offloading when using smartphones with mobile and Wi-Fi networks. Some studies have also examined user characteristics with respect to data traffic offload. Nevertheless, how user behavioural patterns relate to amounts of data traffic offload is poorly understood. The present study addresses this question by segmenting users according to their level of data traffic offload. This work relied on a mixture of data from measurements taken by an application integrated into the user's mobile device, as well as data from a user survey.
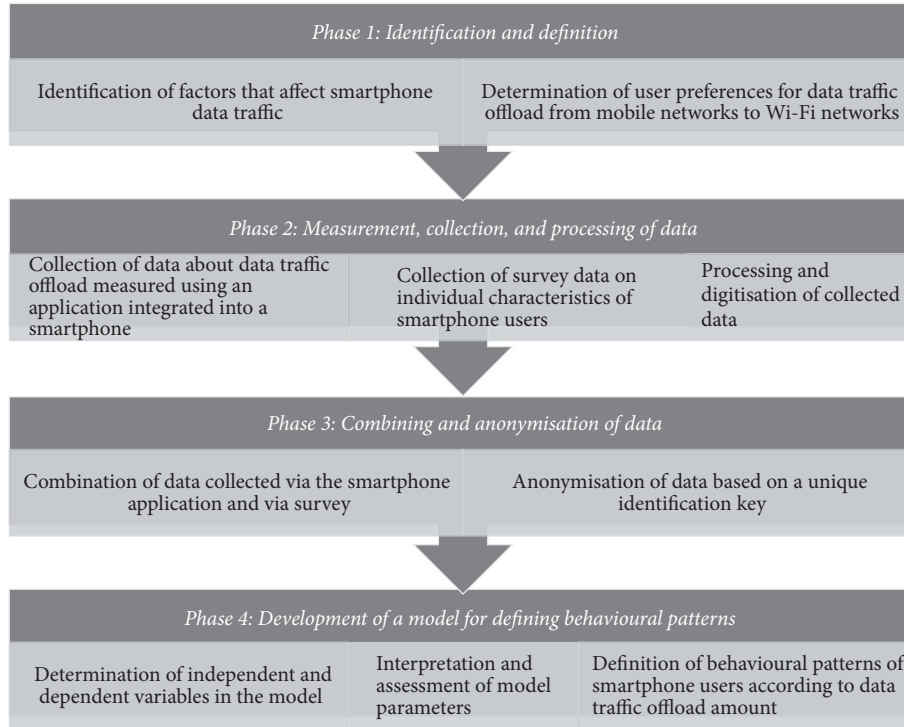
Phase 1: Identification and definition	
Identification of factors that affect smartphone data traffic	Determination of user preferences for data traffic offload from mobile networks to Wi-Fi networks

Phase 2: Measurement, collection, and processing of data		
Collection of data about data traffic offload measured using an application integrated into a smartphone	Collection of survey data on individual characteristics of smartphone users	Processing and digitisation of collected data

Phase 3: Combining and anonymisation of data	
Combination of data collected via the smartphone application and via survey	Anonymisation of data based on a unique identification key

Phase 4: Development of a model for defining behavioural patterns		
Determination of independent and dependent variables in the model	Interpretation and assessment of model parameters	Definition of behavioural patterns of smartphone users according to data traffic offload amount

FIGURE 1: Methodology for the development of the model in the paper.

3. Research Methodology

The methodology for defining behavioural patterns of smartphone users according to data traffic offload from mobile to Wi-Fi networks is presented in Figure 1.

3.1. Identification of Factors and Determination of Preferences. The first phase of the model development methodology presented in Figure 1 includes the following:

(a) Identification of relevant factors that affect the generation of smartphone data traffic.

(b) Determination of user preferences for data traffic offload from mobile networks to Wi-Fi networks.

Relevant factors and user preferences identified in the first phase formed the basis for distinguishing user characteristics directly related to the generation and offload of data traffic. Husnjak et al. [18] identified several factors affecting the generation of smartphone data traffic, including communication technology of the mobile network, size and resolution of the device monitor, device operating system, and price plan and options. Those authors then evaluated each factor using an adequate mathematical method, and those results were incorporated into the present work together with user preferences that Husnjak et al. [19] determined to influence user smartphone data traffic offload from mobile to Wi-Fi networks; those preferences include data offload speed of the Wi-Fi network, availability of the Wi-Fi network, and price structure of access to that network. The factors and user preferences identified in these previous two studies served as

the basis for preparing the online survey in the present work (see Section 3.2).

3.2. Measurement, Collection, and Processing of Data. The second phase of the methodology is collection of data about smartphone data traffic offloading and relevant characteristics of the smartphone users. Data were collected from two sources:

(1) An application integrated into user smartphones provided data about data traffic generated on mobile and Wi-Fi networks.

(2) An online survey provided data on relevant characteristics of the smartphone users.

Figure 2 represents these two flows of data collection, together with the subsequent phase of data anonymisation and combination.

Data were collected during three weeks from 298 participants who responded to the online survey and whose smartphones provided data on data traffic offloading. Data were collected from the smartphones using an embedded (integrated) application compatible with all Android operating systems that provided access to historical information on the amounts of data traffic generated. Both sets of data were collected from most participants at a single interview (Figure 2), when they were first asked to fill out a paper-based form about data traffic based on the data provided by the integrated application. Then, participants received a link to an online survey, which they completed on their smartphone. Some participants filled in the survey later using a computer or any other electronic device.
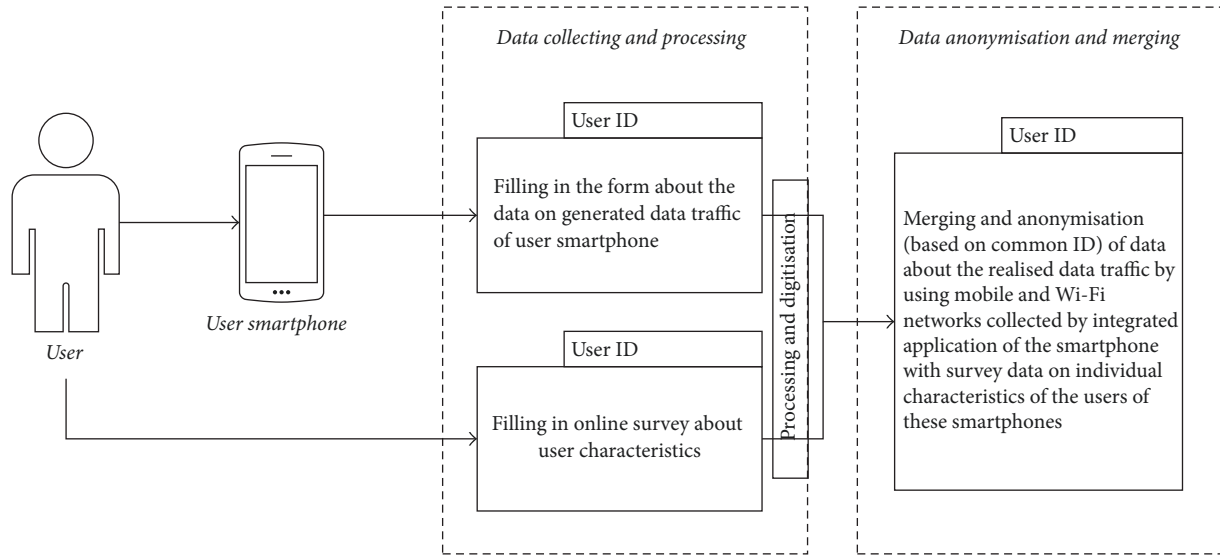
FIGURE 2: Flow of data collection and processing.

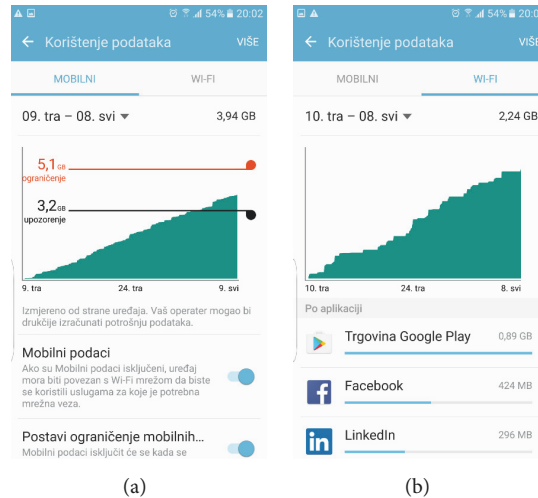(a)   (b)

FIGURE 3: Graphical interface of the integrated application used to measure smartphone data traffic on the Android OS via (a) mobile networks and (b) Wi-Fi networks. The graphical interface shown here corresponds to Android version 6.0.1.

The integrated application provides an overview of the amount of data traffic generated via mobile and Wi-Fi networks (Figure 3), as well as an overview of the data traffic generated by specific applications. Although the graphical interface appeared different based on the Android version, the functionalities were the same. Data traffic was measured over a 28-day period.

Questions on the survey, which were adapted from our previous work [18, 19], took into account various contexts and ecosystems of smartphone use, as well as user behavioural patterns (Table 1). The survey included questions such as "What is the size of the screen on your smartphone?," "How much data traffic is included in your mobile plan?," and "Do you have access to a Wi-Fi network at home?"

3.3. Combining and Anonymisation of Data. The anonymity of data was guaranteed by a unique identification key issued

to each user. This identifier was on the form filled out by the user based on the data measured by the integrated application. It was also the identifier linked to the online survey filled out by the user (Figure 2).

4. Definition of Model Variables

In this paper, the dependent variable was the categorical (ordinal) variable defined according to the amount of smartphone data traffic offloaded from mobile to Wi-Fi networks. The elements and types of variables necessary for the model development are presented in Table 2.

After defining these variables, a database was formed [41] by entering numerical values as their absolute value and entering categorical variables as codes. For instance, the attribute "gender" can be coded such that 1 denotes male and 2 denotes female. In the present study, the dependent

TABLE 1: Previously identified factors and user preferences [18, 19] that served as the basis for the online survey in the present work.

Relevant factors	User preferences
Size and resolution of smartphone display	Wi-Fi network speed
Mobile device operating system	Unlimited data traffic
Device and app settings	Low-cost access
Device capabilities depending on the mobile network	Quality of service/experience
Additional possibilities of applications	Wi-Fi availability in user devices
Mobile network communication technology	Widely deployed infrastructure
Software and applications updates	Quality of mobile network signal
Offload communication technologies	Automated login
Price plan	Security
Use of services and applications	Lower battery drain (when close to access point)
Context of use	International roaming
User personality profile and use of multiple devices	Improved indoor coverage

TABLE 2: Elements of models and types of variables applied to the model in this paper.

Dependent variables (set of alternatives)	Independent variables (set of attributes)
Defined categories of users determined according to the amount of smartphone data traffic offloaded from mobile to Wi-Fi networks	Defined characteristics of users that are expected to affect the selection of the alternative (dependent variable); identification of relevant factors and determination of user preferences allow determination of the set of attributes

variable was categorised according to the amount of data traffic offloaded. Independent variables were taken from the online survey. Both types of variables were brought together to develop a model as depicted in Table 3.

4.1. Determining the Dependent Variables in the Model. In a logit model, independent variables can be continuous or categorical, but the dependent variable is always categorical. Previous authors have pointed out that categorising smartphone users according to data traffic varies from market to market and often reflects the availability and types of price plans [42].

Unfortunately, we could not define our user categories simply based on price plans or on previous studies because of the extensive differentiation of price plans in the Croatian telecommunication market in terms of included data traffic, and also because the inclusion of data traffic in a given plan relates only to data traffic over the mobile network but not over Wi-Fi networks, which is precisely the focus of the present study. Given the lack of consensus among researchers about how to categorise smartphone users according to data traffic offloading, we decided to categorise users into five types (Table 4). Five is often the number of groups chosen in user segmentation research [e.g., [42]]. Additionally, according to [41], the user segmentation can be defined depending on the needs of the researcher. The values of the amount of data traffic offloading (Table 4) are by no means comparable to other research since the authors did not find a research that analyses the smartphone data traffic offloading from mobile to Wi-Fi networks and defines user categories.

These categories were used to classify all users in our sample (Table 5). The total data traffic offloading in the sample varied from 0 to 215 GB.

4.2. Determining the Independent Variables in the Model. The possible independent variables were user characteristics derived from the online survey (Table 6).

As in linear regression, logistic regression involves generating potential models and selecting the one that best explains the given data. The model should have as few variables as possible, since increasing the number of variables increases the standard errors of parameters, rendering the model numerically unstable. One-dimensional analysis followed by multidimensional analysis of the coefficients of independent variables was used to define the best model. Independent variables associated with *p-test* ($p > |z|$) < 0.25 in the one-dimensional analysis were retained in the multidimensional analysis. The final model contained six independent variables (Table 7).

5. Definition of Behavioural Patterns by Determining Probabilities of Event Occurrence

Logistic regression plans the result *(Score) S*, which represents the linear function of the independent variables and related coefficients of logistic regression for single user j:

$$S_j = \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj}, \tag{1}$$

where

S_j is the score for one selected user j;

$\beta_0, \beta_1, \ldots, \beta_k$ are the values of coefficients along with the analysed independent variables;

$x_{1j}, x_{2j}, \ldots, x_{kj}$ are the analysed independent variables for single user j.

TABLE 3: Presentation of the elements of the base of collected data for the development of the model and examples of possible values of variables.

| User ID | Defining of dependent variables (data traffic) | | Defining of independent variables (user characteristics) | | | |
	Amount of data traffic offloaded	Gender	Wi-Fi access	Use of Wi-Fi networks	Other
100001	1 GB	M	Yes	Never	⋯
100002	2 GB	F	No	Sometimes	⋯
100003	3 GB	F	Yes	Always	⋯
⋯	⋯	⋯	⋯	⋯	⋯
	Data from application		Data from survey		

TABLE 4: User segmentation based on data traffic offload.

User category	Data traffic offload
Light user	<1 GB
Medium user	1 GB–5 GB
Medium-heavy user	5 GB–10 GB
Heavy user	10 GB–30 GB
Extreme user	≥30 GB

TABLE 5: Classification of users into predefined categories based on data traffic offload.

User ID	Amount of data traffic offloaded	User category
100001	0.5 GB	Light user
100004	4.3 GB	Medium user
100007	7.2 GB	Medium-heavy user
100008	18.7 GB	Heavy user
100011	56.2 GB	Extreme user
		Dependent variable

Applying Table 7 and scores calculated with the command *ologit* in Stata yielded the following:

$$S_j = -0.496x_{3j} + 0.197x_{9j} - 0.434x_{10j} + 2.379x_{13j}$$
$$- 0.868x_{16j} + 0.331x_{17j}, \quad (2)$$

where

S_j is the result for one defined (selected) user j;

x_{3j}, x_{9j}, x_{10j}, x_{13j}, x_{16j}, x_{17j} are the levels of independent variables determined as significant in the developed model of this paper for single user j.

Consequently, the forecasts of the probability of occurrence of a certain category of the dependent variable of ordinal logistic regression represent the probability that $S + u_j$ lies within the interval between the determined points of intersection (cut) [43, 44]. Here, parameter u_j represents the value of random error. Therefore, the forecast of the probability of occurrence of a certain category of dependent variable for a single user is calculated as follows:

$$P(Y = 0) = P(S_j + u \le \text{cut1}) = \frac{1}{1 + e^{S_j - \text{cut1}}}. \quad (3)$$

This holds for category of users, *light users,* who offloaded the smallest quantity of data traffic from mobile to Wi-Fi networks. The probability of belonging to the next higher category, in this case *medium users,* is

$$P(Y = 1) = P(\text{cut1} < S_j + u \le \text{cut2})$$
$$= \frac{1}{1 + e^{S_j - \text{cut2}}} - \frac{1}{1 + e^{S_j - \text{cut1}}}. \quad (4)$$

In general, the probability of belonging to a certain category of a dependent variable ($P(Y \le j)$) represents the difference of cumulative probabilities, as follows:

$$P(Y = n) = P(Y = n + 1) - P(Y = n). \quad (5)$$

This equation was used to calculate probabilities that a smartphone user with given characteristics belongs to a certain category of dependent variables according to the quantity of data traffic offloaded.

6. Results and Discussion

6.1. Data Analysis. A total of 298 participants filled out the form requesting data traffic generation based on measurements made by the integrated application on their smartphones, and they completed the online survey about user preferences. Of this total, 221 (74.16%) were men and 77 (25.84%) were women. According to the survey, 98.9% of participants use Wi-Fi networks, and 74.16% reported using Wi-Fi networks most often when accessing the Internet and when up- or downloading data. Most participants (77.85%) reported using only Wi-Fi networks when updating applications on their smartphones, while 14.43% reported using both mobile and Wi-Fi networks to do so.

An example of 28-day history of data traffic via mobile and Wi-Fi networks for a single study participant is shown in Table 8. For the purposes of the present study, the amount of data traffic generated via Wi-Fi networks was taken to be the same as the amount of data traffic offloaded from mobile to Wi-Fi networks. This is because the main focus of this study was data offloading.

Table 8 shows a large difference in data traffic between mobile and Wi-Fi networks, which was the case for our study population overall (Figure 4). This graph shows that

TABLE 6: Database formed from user characteristics.

User ID	Gender	Wi-Fi access	Use of Wi-Fi networks	Other
100001	M	Yes	Never	$\cdots$
100002	F	No	Sometimes	$\cdots$
100003	F	Yes	Always	$\cdots$
100004	F	No	Never	$\cdots$
100005	M	Yes	Sometimes	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

Selection of independent variables

TABLE 7: Independent variables that emerged as significant from multidimensional analysis.

Independent variable identification and description		Levels of independent variables (answers to survey questions)
$x3$	Earning income	0: No 1: Yes
$x9$	Quantity (GB) of mobile Internet included in price plan	0: <1 GB 1: 1-2 GB 2: 2-3 GB 3: >3 GB 4: I do not know
$x10$	Is the quantity of GB mobile internet included in the price plan sufficient?	0: Never 1: Sometimes 2: Always
$x13$	Wi-Fi access at home/house?	0: No 1: Yes
$x16$	Application updating	0: Exclusively Wi-Fi 1: Both mobile and Wi-Fi 2: Exclusively mobile
$x17$	Application setting to "use only on Wi-Fi networks"	0: Never 1: Sometimes 2: Always

users conducted nearly 7.7-fold more data traffic over Wi-Fi networks than over mobile networks.

The data traffic offloaded onto Wi-Fi networks by all 298 users involved a total of 564 applications used on Wi-Fi networks (Table 9).

Analysis of individual application data in Figure 4 and Table 9 shows that YouTube was associated with the largest amount of data offloaded to Wi-Fi (1151.85 GB), followed by Facebook (450.51 GB) and supplementary apps (341.48 GB). This last category includes myriad applications used by the user, such as applications for weather forecasting, health/fitness, and web searching. The music streaming application Deezer was associated with the smallest amount of offloaded data (14.94 GB).

Analysis of application categories shows that the categories associated with the largest amounts of offloaded data were *video players and editors* (1182.50 GB), *social* (712.07 GB), and *communication* (348.80 GB). These amounts of offloaded data accounted for 73% of total data traffic offloaded onto Wi-Fi networks, which is consistent with previous studies [5, 42]. The categories associated with the

smallest proportions of data offloading were *other* (80.99%), *social* (82.34%), and *communication* (82.85%). These low proportions of offloaded data likely reflect the importance of these applications to our primarily millennial users for staying in touch with their social contacts and with the latest trending content. In addition, applications such as Facebook and Instagram operate continuously in the background and are accessed occasionally by the user, regardless of whether the Internet network being used is a mobile network or a Wi-Fi network. In contrast, 97.81% of all data traffic in the *entertainment* category occurred via Wi-Fi offload, probably reflecting the notion that users did not consider these applications as constant needs and so chose to use them nearly exclusively when a Wi-Fi network was available.

6.2. Model Analysis. The model was generated with the independent variables $x3$, $x9$, $x10$, $x13$, $x16$, and $x17$, for which the respective coefficients (β) of ordinal logistic regression are shown in Table 10.

The coefficient β measures the direction and magnitude of the change in the log likelihood that the user would

TABLE 8: Example of 28-day data traffic for a single smartphone user in our study.

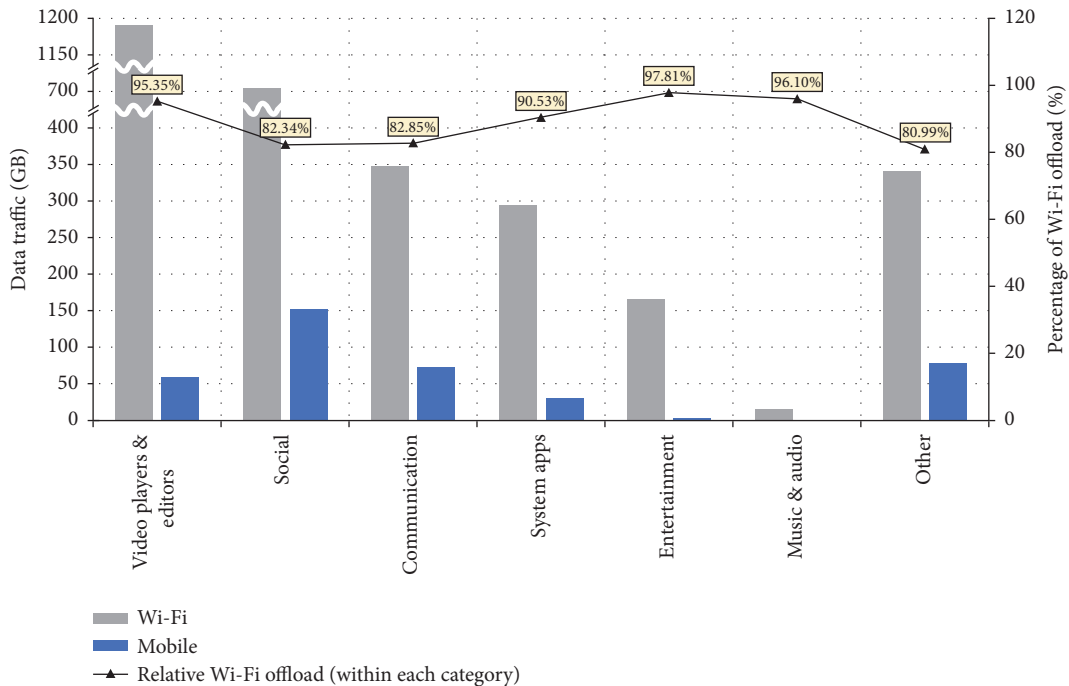| ID	Mobile networks			Wi-Fi networks		
	Application	Amount		Application	Amount	
	Facebook	507	MB	Android OS	1.31	GB
	Opera	300	MB	Facebook	1.31	GB
	YouTube	29.8	MB	Google Play Store	569	MB
	Google Play Store	28.01	MB	YouTube	396	MB
	WhatsApp	27.45	MB	Messenger	245	MB
	Maps	12.41	MB	Opera	221	MB
	Večernji list [Croatian national newspaper]	9.75	MB	WhatsApp	25.55	MB
1062014	Google Play Services	7.78	MB	Gmail	19.49	MB
	Messenger	5.56	MB	Večernji list [Croatian national newspaper]	11.99	MB
	Android OS	5.06	MB	Google Play Services	6.41	MB
	Weather clock	1.76	MB	Video player	5.24	MB
	Skype	1.48	MB	Maps	2.6	MB
	Flipboard	1.44	MB	Google applications	2.3	MB
	Gmail	1.03	MB	Skype	1.02	MB
	Samsung Push Service	779	KB	Private Photo Vault	798	KB
28 days	0.92		GB	4.11		GB

FIGURE 4: Amount of data traffic offloading from mobile to Wi-Fi networks for different categories of applications (as defined in Google Play Store) for all 298 users, based on data provided by the integrated application on users' smartphones.

fall into a lower or higher category of data offloading following a 1-unit increase in the independent variable. Thus, the independent variables $x3$, $x10$, and $x16$ reduced the likelihood that the user would fall into a higher data offloading category, while $x9$, $x13$, and $x17$ increased this likelihood.

To use the model to calculate probabilities that a given user would fall into each of the offloading categories, concrete values of independent variables (such as for the example shown in Table 11) were entered into (1). For example, the probability that the individual in Table 11 is a *light user* ($Y = 0$) is

TABLE 9: Categories and types of smartphone applications generating the highest levels of data traffic offloaded from mobile to Wi-Fi networks, based on data from all 298 users.

Application category*	Application	Data traffic and offloading (within each application)			Distribution of data traffic (within each category)		Distribution of data traffic (across all categories)	
		Wi-Fi offload (GB)	Mobile (GB)	Wi-Fi offload (%)	Relative Wi-Fi offload (%)	Relative mobile data traffic (%)	Relative Wi-Fi offload (%)	Relative mobile data traffic (%)
Video players & editors	YouTube	1151.85	57.37	95.26	95.35	4.65	38.64	14.49
	μTorrent	18.22	0.01	99.96				
	Video player	12.43	0.29	97.69				
Social	Facebook	450.51	94.91	82.60	82.34	17.66	23.27	38.39
	Instagram	227.15	53.77	80.86				
	Snapchat	34.42	4.08	89.40				
Communication	Chrome	235.45	53.15	81.58	82.85	17.15	11.40	18.15
	Internet	59.64	9.61	86.13				
	WhatsApp	34.79	5.15	87.12				
	Browser	18.92	4.32	81.42				
System apps	Google Play Store	195.34	13.05	93.74	90.53	9.47	9.62	7.74
	Removed apps	39.50	9.34	80.87				
	Android OS	33.83	5.09	86.93				
	Google Play	12.99	0.78	94.32				
	Google services	12.84	2.55	83.46				
Entertainment	Showbox	76.78	0.67	99.13	97.81	2.19	5.43	0.93
	Twitch	60.77	0.58	99.05				
	9GAG	15.82	2.46	86.53				
	Netflix	12.71	0.00	100				
Music & audio	Deezer	14.94	0.61	96.10	96.10	3.90	0.49	0.15
Other	Supplementary apps	341.48	80.13	80.99	80.99	19.01	11.16	20.14

* As defined in Google Play Store.

Table 10: Significant independent variables and values of logistic coefficients in the model.

Independent variable	Coefficient value (β)
$x3$	-0.4964411
$x9$	0.1974707
$x10$	-0.4343987
$x13$	2.379751
$x16$	-0.868442
$x17$	0.331091

$$P(Y = 0) = P\left(S_j + u \leq \mathrm{cut}1\right) = \frac{1}{1 + e^{S_j - \mathrm{cut}1}}, \tag{6}$$

where S_j is defined in expression (2). From this, it follows that

$$S_{1071194} = -0.496 \cdot 1 + 0.197 \cdot 1 - 0.434 \cdot 2 + 2.379 \cdot 0$$
$$- 0.868 \cdot 1 + 0.331 \cdot 0, \tag{7}$$
$$S_{1071194} = -2.035,$$

where

> $j = 1071194$ (unique identifier of the selected user, unique user).

The value cut 1 in expression (6) represents the value of the first point of intersection, which is determined by the Stata software. The number of cut values is always one smaller than the number of dependent variable categories, so there were four cut values for the five categories of dependent variables in our study. For the user in Table 11,

$$P(Y = 0) = \frac{1}{1 + e^{-2.035 - (-0.8932006)}} = \frac{1}{1 + e^{-1.1417994}}$$
$$= \frac{1}{1.319244}, \tag{8}$$
$$P(Y = 0) = 0.7582.$$

In other words, the probability that the user in Table 11 is a *light user* for data traffic offloading from mobile to Wi-Fi networks is 75.82%. The probability that the same user is a *medium user* ($Y = 1$) is

$$P(Y = 1) = P\left(\mathrm{cut}1 < S_j + u \leq \mathrm{cut}2\right)$$
$$= \frac{1}{1 + e^{S_j - \mathrm{cut}2}} - \frac{1}{1 + e^{S_j - \mathrm{cut}1}}, \tag{9}$$

where

$$S_{1071194} = -0.496 \cdot 1 + 0.197 \cdot 1 - 0.434 \cdot 2 + 2.379 \cdot 0$$
$$- 0.868 \cdot 1 + 0.331 \cdot 0, \tag{10}$$
$$S_{1071194} = -2.035,$$

so that

$$P(Y = 1) = P\left(\mathrm{cut}1 < S_j + u \leq \mathrm{cut}2\right)$$
$$= \frac{1}{1 + e^{-2.035 - 1.208}} - \frac{1}{1 + e^{-2.035 - (-0.893)}},$$
$$P(Y = 1) = \frac{1}{1 + e^{-2.035 - 1.208}} - \frac{1}{1 + e^{-2.035 - (-0.893)}} \tag{11}$$
$$= \frac{1}{1.039046} - \frac{1}{1.319180},$$
$$P(Y = 1) = 0.2042.$$

This process can be repeated for all the categories (Table 12).

This approach can be used to explore what behavioural patterns are consistent with high probabilities of belonging to certain categories of data traffic offloading, as illustrated in the subset of study participants shown in Table 13. This analysis for the entire study sample is shown in Annex 1.

Analysis of the subset of participants in Table 13 indicates certain trends; for example, people who have levels of independent variables ($x3 = 1$, $x9 = 2$, $x10 = 2$, $x13 = 0$, $x16 = 1$, and $x17 = 0$) are most likely (72.02%) to be in category 0, while people who have levels of independent variables ($x3 = 0$, $x9 = 3$, $x10 = 1$, $x13 = 0$, $x16 = 0$, and $x17 = 1$) are most likely (47.17%) to be in category 1.

Further analysis of individual coefficient values suggests factors that may help classify users according to their offloading behaviour. For example, users with their own income during studies ($x3$) are less likely to belong to a higher category of data offloading than users without their own income. This is understandable, since those with their own income are less concerned with offloading because they can afford to purchase mobile plans that include a larger data transfer amount. Consistent with this idea, users with sufficient data transfer allowance in their mobile plan ($x10$) are less likely to belong to a higher data offloading category. Users who update their applications exclusively on Wi-Fi networks ($x16$) are nevertheless willing to transfer data over mobile networks as well, which is reflected in their lower likelihood of falling into a higher offloading category.

Users who have Wi-Fi access at home ($x13$) are much more likely to fall into higher offloading categories, as are users who define their application settings to operate only on Wi-Fi networks ($x17$). Interestingly, users with larger data allowances in their mobile plans ($x9$) are more likely to fall into higher data offloading categories than users with smaller data allowances. This may reflect the fact that most users with larger data allowances generate a relatively large amount of data traffic, regardless of the network involved.

Useful insights could be gained by incorporating additional information into the model, such as the timing of data offloading and at what types of Wi-Fi access points users offloaded their data, according to a particular application installed on smartphone. As a necessary requirement to this, a third-party application that provides measurements of such and similar phenomena is imposed, since the possibilities of integrated application used in this research are utilised in the best possible way. This falls outside the scope of the present

TABLE 11: Example of one smartphone user's values for the independent variables in the model.

Independent variable	Level of independent variable	Description of independent variable
$x3$: Earning income	1	Yes
$x9$: Quantity (GB) of mobile Internet included in the price plan	1	1-2 GB
$x10$: Is the quantity (GB) of mobile Internet included in the price plan sufficient?	2	Sometimes
$x13$: Wi-Fi access at home/house?	0	No
$x16$: Application updating	1	Both mobile and Wi-Fi
$x17$: Application setting to "use only on Wi-Fi networks"	0	Never

TABLE 12: Probabilities that the individual in Table 11 belongs to each of the categories of data traffic offloading.

Levels of independent variables						Probability of belonging to a certain category of dependent variables				
$x3$	$x9$	$x10$	$x13$	$x16$	$x17$	0	1	2	3	4
1	1	2	0	1	0	75.82%	20.42%	2.75%	0.87%	0.13%

TABLE 13: Subset of study participants showing various behavioural patterns and corresponding variation in the probability of belonging to each category of data traffic offloading.

ID	Levels of independent variables						Probability of belonging to a certain category of dependent variables				
	$x3$	$x9$	$x10$	$x13$	$x16$	$x17$	0	1	2	3	4
1071194	1	1	2	0	1	0	75.82%	20.42%	2.75%	0.87%	0.13%
1037942	1	2	2	0	1	0	72.02%	23.44%	3.32%	1.06%	0.15%
1074764	0	0	2	0	1	0	69.93%	25.07%	3.65%	1.17%	0.17%
1047938	0	0	2	0	1	0	69.93%	25.07%	3.65%	1.17%	0.17%
1080884	1	1	2	0	1	1	69.25%	25.60%	3.76%	1.21%	0.18%
1052018	1	1	2	0	1	1	69.25%	25.60%	3.76%	1.21%	0.18%
1044980	1	3	2	0	1	1	60.28%	32.26%	5.41%	1.79%	0.26%
1055282	0	1	2	0	1	1	57.82%	33.99%	5.93%	1.97%	0.29%
1056608	1	1	2	0	0	1	48.59%	39.95%	8.22%	2.83%	0.42%
1067012	1	3	0	0	1	0	46.99%	40.89%	8.68%	3.01%	0.44%
1060280	0	3	1	0	1	0	45.44%	41.75%	9.14%	3.19%	0.47%
1085474	0	1	0	0	1	0	44.47%	42.28%	9.44%	3.31%	0.49%
1082312	0	1	2	0	0	1	36.52%	45.95%	12.31%	4.54%	0.68%
1055384	1	2	2	1	2	0	36.22%	46.06%	12.44%	4.59%	0.69%
1048448	1	0	1	0	0	2	34.87%	46.53%	13.01%	4.85%	0.73%
1088534	0	2	2	0	0	1	32.07%	47.35%	14.29%	5.46%	0.83%
1087208	0	0	1	0	0	1	31.22%	47.55%	14.70%	5.66%	0.86%
1001426	1	3	1	0	0	1	29.20%	47.93%	15.74%	6.18%	0.95%
1087310	0	3	0	0	1	1	27.92%	48.08%	16.44%	6.55%	1.01%
1084964	0	0	2	1	2	1	26.93%	48.15%	17.00%	6.86%	1.06%
1001120	1	0	2	1	1	0	26.13%	48.18%	17.47%	7.11%	1.11%
1055690	1	0	2	1	1	0	26.13%	48.18%	17.47%	7.11%	1.11%
1050386	1	0	0	0	0	2	25.75%	48.18%	17.70%	7.24%	1.13%
1075070	1	2	0	0	0	1	24.55%	48.13%	18.45%	7.67%	1.20%
1060178	0	0	0	0	0	1	22.72%	47.90%	19.65%	8.40%	1.33%
1001630	1	1	2	1	1	0	22.50%	47.86%	19.80%	8.49%	1.35%
1063646	1	1	2	1	1	0	22.50%	47.86%	19.80%	8.49%	1.35%
1061300	1	1	2	1	1	0	22.50%	47.86%	19.80%	8.49%	1.35%
1078844	1	1	2	1	1	0	22.50%	47.86%	19.80%	8.49%	1.35%
1078436	1	1	2	1	1	0	22.50%	47.86%	19.80%	8.49%	1.35%
1076804	1	3	0	0	0	1	21.08%	47.51%	20.81%	9.15%	1.46%
1089656	1	0	2	1	1	1	20.26%	47.24%	21.41%	9.56%	1.53%
1088330	0	3	1	0	0	1	20.06%	47.17%	21.56%	9.66%	1.55%

study, which focused on the volume of data offloaded from mobile to Wi-Fi networks.

7. Conclusions

This study has demonstrated the possibility of using ordinal logistic regression to correlate user preferences or characteristics with the likelihood of offloading different amounts of mobile data from mobile to Wi-Fi networks. This model provides one of the most extensive analyses of smartphone user behavioural patterns and how that behaviour may relate to data offloading amount. This model, and the approach of collecting data from user surveys as well as from applications installed on their smartphones, may be useful for mobile network operators and service providers for understanding customer behaviour, segmenting their markets, and differentiating their price plans. In addition, the results and approach may be useful for Wi-Fi network operators and service providers, who control an ever-growing share of data traffic but who currently lack market analytics to analyse, for example, the increasingly important "blackhole" of data traffic offloading. The modeling of offloading described here will help various stakeholders better exploit the revenue potential of the offloading market.

This model can help network operators and service providers offer special options or differentiated price plans to take advantage of the tendency of certain users to use Wi-Fi networks. For instance, operators may offer unlimited *e-mail* for *roaming* services or a data package that includes free access to popular social networks. The model presented here creates new potential for strategic action, rational optimisation of services and price plans, and personalisation of services that can explicitly take into account the increasingly important Wi-Fi network ecosystem.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Supplementary Materials

This Supplementary Material contains Annex 1 which shows the behavioural patterns and corresponding variation in the probability of belonging to each category of data traffic offloading for the entire study sample. *(Supplementary Materials)*

References

[1] S. Taylor, A. Young, and A. Noronha, *What Do Consumers Want from Wi-Fi? Insights from Cisco IBSG Consumer Research*, Cisco Internet Bussines Solutions Group, 2012.

[2] A. Aijaz, H. Aghvami, and M. Amani, "A survey on mobile data offloading: technical and business perspectives," *IEEE Wireless Communications Magazine*, vol. 20, no. 2, pp. 104–112, 2013.

[3] Ericsson, *Mobile Business Trends—November 2015*, Ericsson, Stockholm, Sweden, 2015.

[4] E. Yaacoub, H. Ghazzai, M.-S. Alouini, and A. Abu-Dayya, "On the impact of D2D traffic offloading on energy efficiency in green LTE-A HetNets," *Wireless Communications and Mobile Computing*, vol. 15, no. 6, pp. 1089–1105, 2015.

[5] Cisco, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020*, Cisco, San Francisco, Calif, USA, 2016.

[6] J. S. Marcus and J. Burns, *Study on Impact of Traffic Off-Loading and Related Technological Trends on the Demand for Wireless Broadband Spectrum*, European Commission DG Communications Networks, Content & Technology, 2013.

[7] K. Teshager, *Strategic Use of Wi-Fi in Mobile Broadband Networks*, University of Vaasa, 2010.

[8] A. Aijaz, N. Uddin, O. Holland, and A. H. Aghvami, "On Practical Aspects of Mobile Data Offloading to Wi-Fi Networks," Ithaca, NY, USA, 2015.

[9] B. Han, P. Hui, V. S. A. Kumar, M. V. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Transactions on Mobile Computing*, vol. 11, no. 5, pp. 821–834, 2012.

[10] D. H. Hagos, *The Performance of WiFi Offload in LTE Networks*, Luleå University of Technology, 2012.

[11] S. Bae, J. Kim, D. Ban et al., "StreetSense: effect of bus Wi-Fi APs on pedestrian smartphone," in *Proceedings of the ACM Internet Measurement Conference*, pp. 347–353, October 2015.

[12] Informa Telecoms & Media, *Understanding Today's Smartphone User, Demystifying Data Usage Trends on Cellular and Wi-Fi Networks Part 2: An Expanded View by Data Plan Size, OS, Device Type and LTE*, Informa Telecoms & Media, London, UK, 2012.

[13] Wireless Broadband Alliance, *Global Developments in Public Wi-Fi*, Wireless Broadband Alliance, Singapore, 2011.

[14] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: how much can wifi deliver?" *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 536–550, 2013.

[15] A. Salter, *A Carrier Roadmap for Monetizing Next Generation Wi-Fi*, iPass, Redwood Shores, Calif, USA, 2012.

[16] Informa, *Understanding Today's Smartphone User: Demystifying Data Usage Trends on Cellular & Wi-Fi Networks*, 2012.

[17] Ovum, "Smartphone & tablet usage trends & insights 4G LTE and Wi-Fi powering data consumption," 2015.

[18] S. Husnjak, D. Peraković, and I. Cvitić, "Relevant affect factors of smartphone mobile data traffic," *Promet—Traffic & Transportation*, vol. 28, no. 4, pp. 435–444, 2016.

[19] S. Husnjak, I. Forenbacher, D. Peraković, and M. Periša, "Preferences of Smartphone Users in Mobile to Wi-Fi Data Traffic Offload," in *Proceedings of the 34th Symposium on Novel Technologies in Postal and Telecommunication Traffic*, pp. 295–305, 2016.

[20] 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.

[21] T. J. 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, no. 4-5, pp. 357–371, 2013.

[22] T. J. Gerpott, S. Thomas, and M. Weichert, "Personal characteristics and mobile Internet use intensity of consumers with computer-centric communication devices: An exploratory empirical study of iPad and laptop users in Germany," *Telematics and Informatics*, vol. 30, no. 2, pp. 87–99, 2013.

[23] S. Alwahaishi and V. Snášel, "Factors influencing the consumers' adoption of mobile internet," in *Proceedings of the 3rd International Conference on Digital Information and Communication Technology and Its Applications*, pp. 31–39, 2013.

[24] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, "A first look at traffic on smartphones," in *Proceedings of the 10th Internet Measurement Conference (IMC '10)*, pp. 281–287, Melbourne, Australia, November 2010.

[25] S. K. Baghel, K. Keshav, and V. R. Manepalli, "An investigation into traffic analysis for diverse data applications on smartphones," in *Proceedings of the 18th National Conference on Communications (NCC '12)*, February 2012.

[26] A. Ghose and S. P. Han, "An empirical analysis of user content generation and usage behavior on the mobile Internet," *Management Science*, vol. 57, no. 9, pp. 1671–1691, 2011.

[27] 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," in *NETWORKING 2012*, vol. 7289 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 318–330, Springer, Berlin, Germany, 2012.

[28] J. Yang, Y. Qiao, X. Zhang, H. He, F. Liu, and G. Cheng, "Characterizing user behavior in mobile internet," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 95–106, 2015.

[29] J. Binde and M. Fuksa, "Mobile technologies and services development impact on mobile internet usage in Latvia," *Procedia Computer Science*, vol. 26, pp. 41–50, 2013.

[30] A. Kivi, "Measuring mobile user behavior and service usage: methods, measurement points, and future outlook," *International Journal of Mobile Communications*, vol. 7, no. 4, pp. 415–435, 2009.

[31] T. Soikkeli, *The effect of context on smartphone usage sessions [M.S. thesis]*, Aalto University School of Science, Aalto, Finland, 2011.

[32] H. T. Verkasalo, *Handset-based analysis of mobile service usage [Ph.D. thesis]*, University of Technology, Helsinki, Finland, 2009.

[33] J. Lee and P. Seeling, "An overview of mobile device network traffic and network interface usage patterns," in *Proceedings of the IEEE International Conference on Electro/Information Technology (EIT '13)*, pp. 1–4, May 2013.

[34] K. Fukuda, H. Asai, and K. Nagami, "Tracking the evolution and diversity in network usage of smartphones," in *Proceedings of the ACM Internet Measurement Conference (IMC '15)*, pp. 253–266, Tokyo, Japan, October 2015.

[35] M. Paolini, *Taking Wi-Fi Beyond Offload: Integrated Wi-Fi Access Can Differentiate Service and Generate New Revenues*, Openet, Dublin, Ireland, 2013.

[36] K. Bakhit, C. Chalouhi, S. Francis et al., "3G to Wi-Fi offloading on Android," in *Proceedings of the IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA '14)*, pp. 247–252, Doha, Qatar, November 2014.

[37] C. Hetting, "Seamless Wi-Fi offload: a business opportunity today," 2013.

[38] S. Deng, R. Netravali, A. Sivaraman, and H. Balakrishnan, "WiFi, LTE, or both? Measuring multi-homed wireless internet performance," in *Proceedings of the ACM Internet Measurement Conference (IMC '14)*, pp. 181–194, Vancouver, Canada, November 2014.

[39] S. Kaisar, *Smartphone Traffic Characteristics and Context Dependencies*, University of Saskatchewan, 2012.

[40] 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.

[41] J. R. Šimićević, *A Contribution to Defining the Parking Policy Depending on the Level of Service in the Traffic Network*, University of Belgrade, Faculty of Transport and Traffic Engineering, Belgrade, Serbia, 2013.

[42] Ericsson, *Ericsson Mobility Report (June 2015)*, Ericsson, Stockholm, Sweden, 2015.

[43] X. Liu, *Applied Ordinal Logistic Regression Using Stata*, SAGE Publications Ltd, Thousand Oaks, Calif, USA, 2016.

[44] O. Torres-Reyna, *Getting Started in Logit and Ordered Logit Regression*, Princeton University, 2012.