Apps on the Move: A Fine-Grained Analysis of Usage Behavior of Mobile Apps

Lin Yang[*], Mingxuan Yuan[†], Wei Wang[‡*], Qian Zhang[*], Jia Zeng[§†]
[*]Hong Kong University of Science and Technology,
[†]Huawei Noah's Ark Lab, Hong Kong,
[‡]School of Electronic Information and Communications, Huazhong University of Science and Technology,
[§]School of Computer Science and Technology, Soochow University, Suzhou

***Abstract*— Owing to the proliferation of mobile devices and their corresponding app ecosystems, more and more people are accessing the internet via various mobile apps, which generates tremendous volume of mobile data. Despite the growing importance of these mobile apps, we have a rather sparse understanding of how they are accessed and what issues affect their usage patterns. To address this problem, we perform a comprehensive measurement on large-scale anonymized network data collected from a tier-1 cellular carrier in China. In this measurement, we characterize the usage pattern of mobile apps and exhibit how the mobility, geospatial properties and behaviours of subscribers affect their mobile app usage at a fine-grained level.**

I. Introduction

Nowadays, accessing the internet via mobile device has become an indispensable part of our daily lives. A recent study [1] reports that global mobile data traffic grew 81 percent in 2013 and reached 1.5 exabytes per month by the end of 2013. Meanwhile, with the convenience brought about by the blooming development of mobile apps and pre-installed marketplace portals, *e.g.*, AppStore on iOS and Google play on android, more and more people are accessing the internet via mobile apps instead of traditional web browsers [2]. In spite of the increasing importance of mobile apps, we only have a sparse understanding of how these apps are used and what critical factors affect their usage patterns, especially at a fine-grained level. To fill this knowledge gap, we perform an in-depth analysis on mobile app usage patterns based on a large-scale data set in this work.

Previous works in this area can be roughly categorized into two groups by their measurement methodology: the first is instrumented measurement, which requires pre-installed apps on volunteer's mobile device to collect detailed log traces. The authors in [3], [4] provide a detailed analysis on app usage patterns and energy consumption based on a small number of logs collected from instrumented phones. Other similar studies are performed in [5], [6]. As the cost of measurement is quite high, these works are generally limited by the scale of the data sets. On the other hand, some studies have performed data-driven measurements based on large-scale data: [7] analyses the geospatial correlation of app interest by using a large-scale data log in a 3G network. Paul *et al.* studied the traffic dynamics from the perspectives of both cellular providers and subscribers in a tier-1 3G network [8]. The authors in [2] provide a study on smartphone app usage patterns from a nation-wide view. However, these works only provide a broad view of the traffic from the network provider's perspective instead of taking an in-depth look at critical issues which affect app usage patterns.

In this work, we collect large-scale anonymized IP flow traces from a tier-1 cellular provider in China, which contains millions of subscribers and covers thousands of cells in one of the biggest metropolitan areas. Based on this data set, we exhaustively investigate how subscriber mobility, geospatial patterns and preference affect the usage pattern of mobile apps.

Our key contributions are summarized as follows:

- Our data set contains information about millions of subscribers, including usage traces of hundreds of apps and covers a large metropolitan area. This rich data set enables us to make statistically meaningful observations about mobile app usage.
- We confirm that there exists a huge diversity in subscribers' mobility. Although most subscribers move within a small area per day, *e.g.*, within 5 cells and a radius of 5 kilometers, there are a substantial number of subscribers who roam across 20 cells or more than 20-kilometer area in a day.
- We observe that the average traffic increases with the subscriber's mobility. This implies the subscribers of higher mobility tend to generate more traffic volume, which motivates our further analysis on the impact of mobility on app usage and behavior of different subscriber groups.
- We also find the impact of mobility varies for each app. For example, the *web browsing* traffic increases with subscribers mobility , and *gaming*, *social networking* apps are more frequently used when subscribers are roaming within a relatively small region. This result suggests OS vendors and service providers of these apps should consider techniques which compensate for network quality variations caused by mobility.
- We validate that there exists a strong correlation between subscriber mobility and the traffic of some apps, *e.g.*, the correlation between the traffic of *maps* apps and the number of cells visited is quite high. This further implies we can estimate subscribers' app traffic with their mobility level.

978-1-4673-9953-1/16/$31.00 ©2016 IEEE

- By identifying hundreds of locations of different function types, we notice that, although location type does not change the possibility of accessing an app, it obviously affects the extent to which mobile apps can be used. For example, *map* generates more traffic in *transportation* areas, and *music* is more highly preferred in *work* areas. This suggests a potential for service providers to optimize their services by placing content servers near the majority of their users. Also, this observation is helpful for network providers as different compositions of app traffic pose various network QoS requirements.
- Our study on behavior of *heavy traffic subscribers* (Top 20% of subscribers in terms of daily traffic) shows several interesting observations: First, heavy traffic subscribers generally have a higher mobility level than normal subscribers (the remaining 80% of subscribers). However, the app usage patterns of heavy traffic subscribers are more sensitive to mobility: their app traffic fluctuates significantly with mobility growth, while the app traffic of other subscribers remains stable. Another observation is that location also affects heavy traffic subscribers more significantly than others. As heavy traffic subscribers contribute to a large fraction of the total traffic, knowing such app usage pattern can be helpful for more efficient network planning and resource allocation.
- We also conduct a comprehensive analysis on app usage patterns of high mobility subscribers (the top 20% subscribers in terms of mobility) and notice that the app interest and sensitivity to location for high mobility subscribers is quite different to other subscribers. Such knowledge could be leveraged by network operators, service providers and app designers to optimize their services case by case.

The rest of this paper is organized as follows: section II describes our data set, section IV investigates the impact of subscriber mobility on app usage patterns, and section V explores the geospatial pattern of app usage. Related works are discussed in section VIII, and then we conclude our work.

II. Data Set

In this work, we use an anonymized data set from a tier-1 cellular network provider in China. This data set contains flow information of more than 8 million subscribers and covers a large metropolitan area of China from September 6th, 2012 to September 18th, 2012. It is collected from all links between Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN) in the core network of a mixed 2G/3G cellular network and contains flow-level information of all the IP flows carried in the PDP context tunnels, that is, flows that are sent to and from mobile devices. The information includes: anonymized subscriber identifiers, the traffic volume of each flow, application information and location information. All subscriber-related identifiers are anonymized to protect privacy without affecting our analysis.

Application information consists of application name (*e.g.*, Google Maps), protocol, IP address and port, delay and transmission speed. Among these fields, the application name is a result of Deep Packet Inspection (DPI), which is conducted by network providers at IP level to improve network security and provide application-specific services [9]. In this work, we use an internal DPI solution provided by the cellular network operator, whose accuracy is more than 90% and sufficient for our study.

The location information for each flow contains Location Area Code (LAC) and Sector ID (SI). This location information is obtained by joining PDP sessions with a fine-grained log of signalling messages, which includes detailed logs of handover events, thus the location should be accurate for each flow.

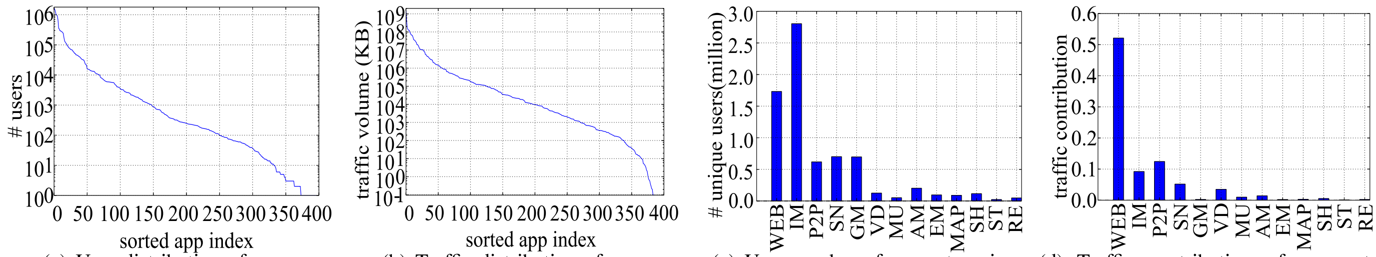
In total, our data set contains more than 8 million users and covers 38360 sectors of a very large metropolitan area. The data record of each day is about 800 GB.

Although there are hundreds of apps in our data set, there exists some apps which are rarely used. After ranking all the apps by the number of average daily users, we notice that some apps are only used by a few users (less than 100) and contribute a very small proportion of the total traffic (less than 10 percent in total). As the usage pattern of these apps could easily be affected by a small group of users and prone to introduce bias, therefore we leave these “tiny” apps out of our analysis and only focus on the top 160 popular apps, which are used by the majority of users and generate a dominant proportion of total data traffic.

Furthermore, we manually categorize these 160 apps into 13 groups according to their function or genre: (1) web browsing(WEB) stands for traditional web browsers, *e.g.*, Chrome, Safari; (2) P2P, *e.g.,* BitTorrent; (3) instant message (IM) includes apps like WeChat or MSN; (4) reading (RE) includes news reading apps like CNN news app, or RSS readers; (5) social network (SN) covers Sina Webo (a popular twitter-like app in China) and Renren(facebook-like app) and other similar apps; (6) video (VD) consists of popular VOD apps in China, like Youku and Sina Video; (7) music (MU) includes popular online music apps in China, *e.g.*, TecentMusic; (8) app market (AM), such as apple app store and android play; (9) game (GM) includes popular online games in China; (10) email (EM) refer to the popular email clients on mobile device, e.g., iOS email app; (11) stock trading (ST) contains stock trading apps, such as apple stocks; (12) online shopping (SH) covers popular online shopping apps in China, like Taobao or Amazon apps; (13) maps, e.g., Google maps or Baidu maps. For each category, we aggregate apps belonging to this category and conduct a fine-grained analysis on its usage pattern in later sections.

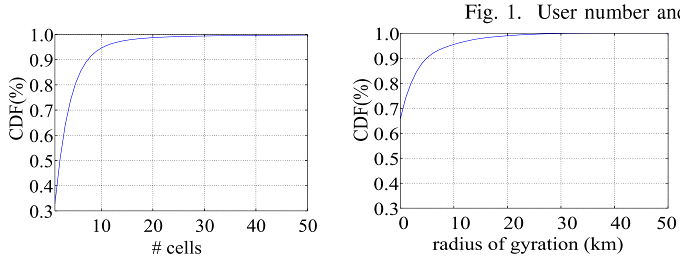
TABLE I
App categories

Category	# apps	Category	# apps
web browsing	6	p2p	9
instant message	6	reading	16
social networks	5	video	9
music	16	app market	3
game	70	email	4
stock trading	8	online shopping	6
map	2		



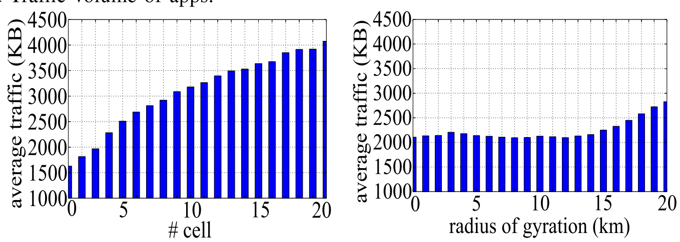
(a) User distribution of apps. (b) Traffic distribution of apps. (c) User number of app categories. (d) Traffic contribution of app categories.

Fig. 1. User number and Traffic volume of apps.



(a) number of visited cells. (b) radius of gyration.

Fig. 2. Subscriber's mobility.



(a) average traffic vs. visited cell (b) average traffic vs. radius of gyration

Fig. 3. Traffic distribution over mobility.

We understand that these apps will be invisible in our analysis once the device is connected to WiFi or other networks, and such situations occur frequently in daily life. However, as mobility is the key advantage of mobile devices and the cellular traffic generated when mobile apps are moving has increased exponentially in recent years, we choose to focus our analysis on the access patterns of mobile apps under the scenario of cellular networks.

III. Overview of App Usage

We begin our analysis by presenting the general overview of all apps, and then manually group those apps into 13 categories according to function or genre and investigate the user and traffic contributions of each category.

We first sort apps according to the user number and traffic respectively and show the user number and traffic distribution of each app per day in Figure 1(a) and Figure 1(b). Note that the y-axis is in a log scale. The long tails of app user numbers and traffic volume distribution imply that there exists a high diversity in app usage: the top app in our data set is a web browsing app based on HTTP, which is accessed by nearly 1.7 million users and contributes more than 1.8 TB of traffic per day, whereas the least popular app is only accessed by 1 user and accounts for a little traffic.

Then, we manually group these apps into 13 app categories by their functions or genres. Table I shows the number of apps in each category. For each category, we aggregate the user number and traffic volume of apps which belong to this category and report the result in the Figures 1(c) and 1(d). We observe that *web browsing* and *instant messaging* are the most popular apps in terms of user number and cover almost all the subscribers. Meanwhile, there are also a substantial number of users who access *game*, *p2p* and *social networks*. However, in terms of traffic volume, *web browsing* dominates: it generates approximately 1.8 TB per day and contributes more than 50% of the total traffic volume of the whole network.

IV. Mobility Pattern

In this section, we examine mobility patterns of subscribers and investigate how subscribers' mobility patterns affect mobile app usage.

A. *Subscriber Mobility*

The first question is what subscribers' mobility looks like. To find the answer to this question, we choose two different metrics to measure subscriber mobility:

(a). Number of visited cells, defined as the number of cells visited along subscriber's daily trajectory. Figure 2(a) shows the distribution of the number of visited cells: 80% of subscribers are only observed within 5 cells per day and 95% of subscribers roam less than 10 cells per day. However, about 3% of subscribers move across more than 20 cells, recall that we have 8 million subscribers in our data set, thus 3% still implies there are a considerable number of subscribers who have very high mobility.

Owing to the fact that the coverage of cell towers varies, ranging from several kilometers to tens of kilometers [10], the *number of visited cells* can not perfectly represent the geospatial coverage of subscriber activities. Therefore, we further study subscriber mobility via other metrics: **(b). Radius of gyration** (RoG), which is commonly used in the study of human mobility [11] and can be interpreted as the geographical area travelled by a subscriber:

$$r_g = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(l_i - l_{mass})^2}, \qquad (1)$$

where l_i is the latitude and longitude of cell i, $l_{mass} = \frac{1}{n}\sum_{i=1}^{n} l_i$ is the center of mass of subscriber's trajectory and $(l_i - l_{mass})$ is the Euclidean distance between l_i and l_{mass}. The *radius of gyration* is a *de facto* indicator of subscribers' activity areas: the higher it is, the larger the area the subscriber travels.

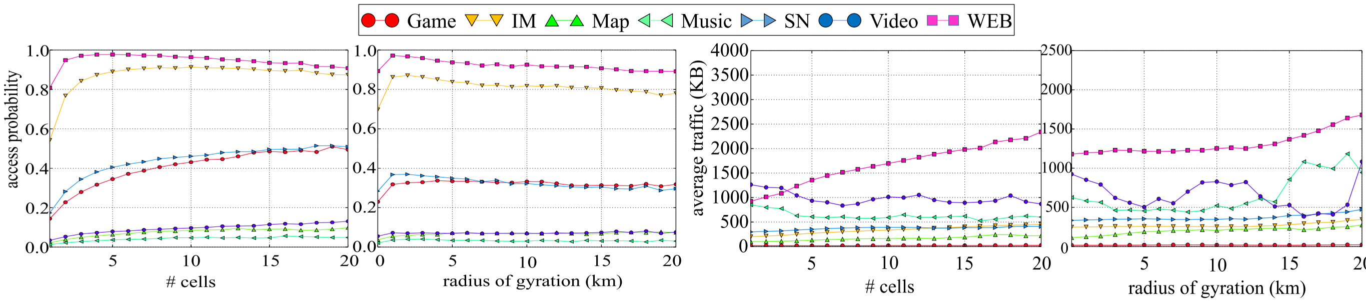
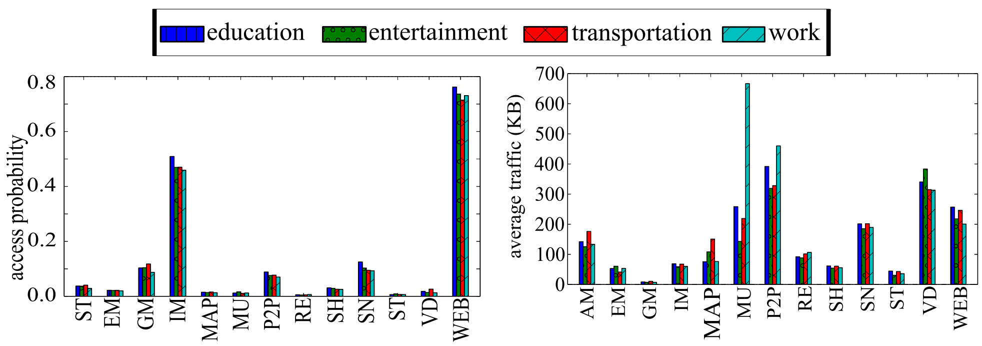


Fig. 4. Impact of mobility on app access probability.

Fig. 5. Impact of mobility on app traffic.



(a) App access probability in cells of different types.

(b) App traffic contribution in cells of different types.

Fig. 6. Impact of location on app usage

We compute the *radius of gyration* of each subscriber every day and report it in Figure 2(b). A similar high diversity of mobility can be observed in this figure: over 90% of subscribers move within a radius of 5 kilometres, while there is still an approximate 2% of subscribers who travel across an area larger than 20 kilometers per day.

High mobility leads to variations in network connectivity and quality, and thus eventually affects app usage. To understand how mobility affects app usage patterns, we study the effect of subscriber's mobility in the next subsection.

B. Impact of Mobility on App Usage

We first examine traffic generated by subscribers of different mobility. Figure 3 shows the distribution of the average daily traffic per subscriber in their mobility. We note there is an increasing trend of average traffic volume when a subscriber's mobility grows in terms of both *number of visited cells* and *radius of gyration*, which implies higher mobility subscribers tend to generate more traffic. Similar observations are also found in [7], [8]. However, since many confounding factors still exists, previous works only describe the aggregated behavior, and have not investigated the root of this phenomenon in detail. In our study, we investigate two of the most important confounding factors: app category and subscriber group.

We start with the impact of mobility on different app categories: a natural question for this factor is, *how does mobility affect usage patterns of apps of different app categories respectively?* To answer this question, we first define the access probability of the ith app under mobility m as:

$$prob(a_{im}) = \frac{u_{im}}{\sum\limits_{i \in S} u_{im}}, \tag{2}$$

where u_{im} is the number of subscribers who access the i-*th* application under mobility m, and S is the set of all apps. Then, we aggregate apps into categories in Figure 4. To highlight variations, some apps whose trends remain stable are omitted in this figure, *e.g., shopping* and *reading*

One obvious trend is that, the access probability of *web browsing*, *instant messager* grows significantly at the beginning and then remain at a high volume as mobility increases. Also, the access probability of *gaming* and *social networks* show a notable growth when the number of visited cells increase. This could be a result of the fact that people tend to stay connected with others when they are out of their comfort zones [7]. However, when using the *radius of gyration* as a mobility indicator, the access probability of each app category is relatively stable, except for a short increase at the beginning. This implies that only subscribers who move within a small region but with a large visited cell number, *e.g.*, commute in downtown) prefer *gaming*, *social networks* than other app categories.

Apart from this, we also examine the traffic contributions of each app category. We first aggregate traffic generated by each app category, then average it on each mobility and report the result in Figure 5.

We notice an interesting phenomenon: the average traffic of *web browsing* and *music* increase significantly with subscribers' mobility. This trend is suspicious, because high mobility would degrade network connectivity and quality [12], such data-sensitive apps should disappear as mobility increases. To find the reason behind this, we trace the subscribers who consume heavy traffic of these apps in high mobility, and surprisingly find that these subscribers always tend to generate

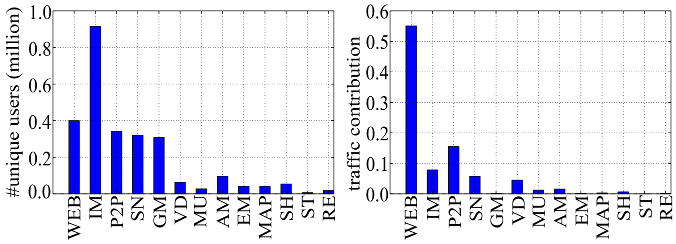


Fig. 7. App interest of heavy traffic subscribers.

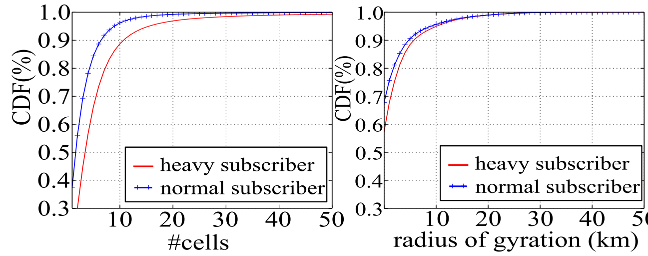
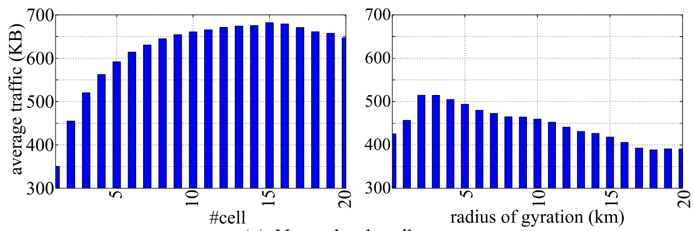
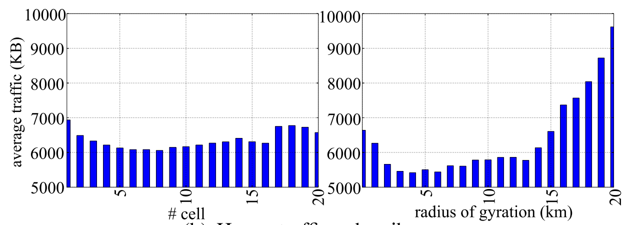


Fig. 8. Mobility of heavy traffic subscribers.



(a) Normal subscribers.



(b) Heavy traffic subscribers.

Fig. 9. Impact of mobility on app traffic for normal/heavy traffic subscribers.

TABLE II

CORRELATION BETWEEN APP TRAFFIC AND MOBILITY

Category	r with #cells	r with RoG
web browsing	0.96	0.15
p2p	0.17	0.04
instant message	0.99	0.11
reading	0.15	-0.32
social networks	0.58	-0.09
video	-0.20	-0.46
music	-0.37	0.37
app market	0.18	-0.22
game	0.13	0.14
email	0.29	0.21
stock trading	-0.42	-0.32
online shopping	-0.53	-0.34
map	0.72	0.69

a large volume of traffic and are generally more mobile than other subscribers. This motivates our further analysis of app usage behavior of heavy traffic subscribers and details are discussed in section VI.

After finding that the impact of mobility on each app varies, we explore whether any connection exists between subscriber mobility and app usage: after aggregating app traffic into categories, the Pearson product-moment correlation coefficient [13] is computed as follows:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X}^2)}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}, \tag{3}$$

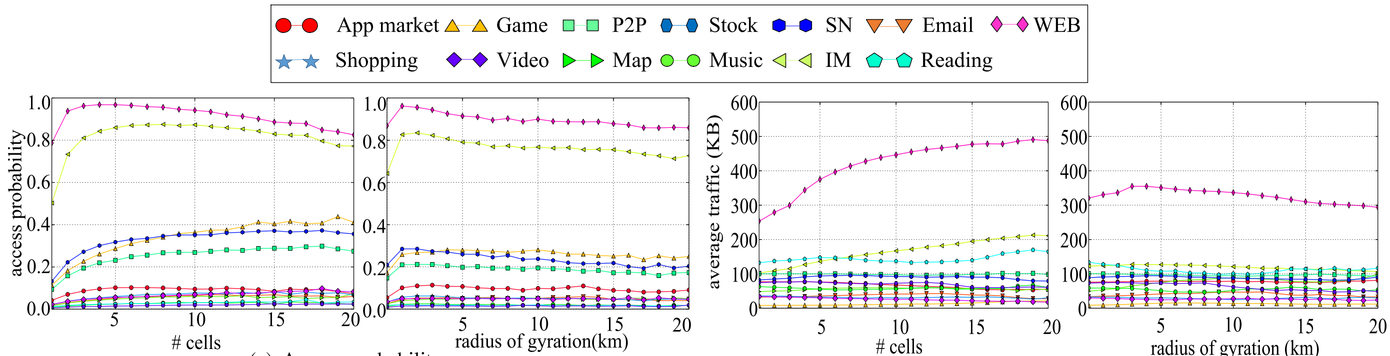
which could be interpreted as a measure of the linear correlation between app traffic volume and mobility. Table II presents the correlation between the average traffic of each app category and mobility. We can see a strong positive correlation between the number of visited cells and the average traffic of some app categories, *e.g.*, *web browsing*, *instant message* and *map*, while a relatively strong negative correlation exists in *stock trading* and *online shopping*. A similar correlation can also be observed between the radius of gyration (RoG) and average traffic of *map*. These high correlations suggest that subscriber's mobility would be helpful when predicting app traffic volume.

V. SPATIAL PATTERN

In this section, we explore the effects of the geographic location on app usage. Understanding such spatial patterns provides the possibility of conducting location-based service optimization, *e.g.*, content providers could benefit in optimizing the placement of their content delivery servers for better service quality.

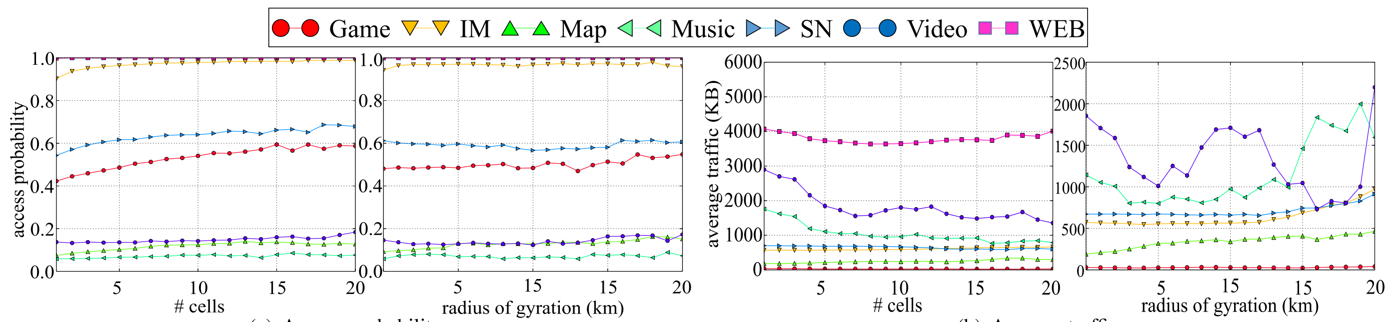
Several previous studies have conducted such measurements at a relatively coarse level, either nation-wide [2] or based on a small number of regions, *e.g.*, downtown and suburban [8], [14]. To take a fine-grained study at the geospatial pattern of app usage, we first estimate the coverage of cells according to the distance between nearby cells, then heuristically classify cells into four types according to their function: **(a) Transportation**, such as train stations or airports; **(b) Educational institution** includes schools, colleges and research institutions; **(c) Work** consists of business districts, campuses of large technology corporations and government office areas; **(d) Entertainment** covers large shopping malls and places of interest in the downtown. Due to page limitations, the detail of classification is not included here. In total, there are 38360 cells in our data set, we have identified the types of 1241 cells in the downtown area and all classification results are validated manually one by one. Table III shows the number of cells of each type.

We understand that there are some compound cells whose types overlap, *e.g.*, in some downtown areas, large shopping malls coexist with office buildings, thus it can be *Work* and *Entertainment* at the same time. Moreover, the proportion of each type for such compound cell is ambiguous, which makes its impact on app usage more difficult to investigate. Therefore, we only examine cells which have a significantly dominant type. Another difficulty arises from the fact that the cell type



(a) Access probability (b) Average traffic.

Fig. 10. Impact of mobility on app usage of normal subscribers.



(a) Access probability (b) Average traffic.

Fig. 11. Impact of mobility on app usage of heavy traffic subscribers.

is semantic, *i.e.*, the same location may have various meanings for different people at different times. Since we are primarily interested in the impact of location type for the majority of subscribers, this is not included in this measurement. Also, since most of people are using WiFi or LAN at home, we do not include areas of *residence* in our analysis.

We first compare the access probability of each app category in cells of different types in Figure 6(a). From this figure, the access probability merely changes in different locations: *instant messaging* and *web browsing* are always the most popular apps in cells of different types, and the proportion of other apps also only make a minor difference. However, this result is out of our expectation that location should play an import role in app usage [2], [7].

TABLE III
CELL TYPES

Type	# cells
transportation	158
education	300
work	299
entertainment	484

To further investigate this problem, we compare the average traffic generated by each subscriber in cells of different types in Figure 6(b) and find the traffic patterns of some apps vary with locations. For instance, *music* only dominates in *work* areas, *P2P* is more frequently used in *education* and *work* areas than other areas, while *map* contributes the highest traffic in *transportation* areas. This demonstrates, instead of bringing significant variation in access probability, location affects the extent to which a mobile app is used. This suggests that, knowing the coverage area types of cells, it is possible to estimate the network quality requirements as their app traffic composition varies with cell types and each app category poses different network quality requirement, *e.g.*, a cell covering *work* areas implies a requirement of high bandwidth such as *music* and *P2P* are frequently used within this area.

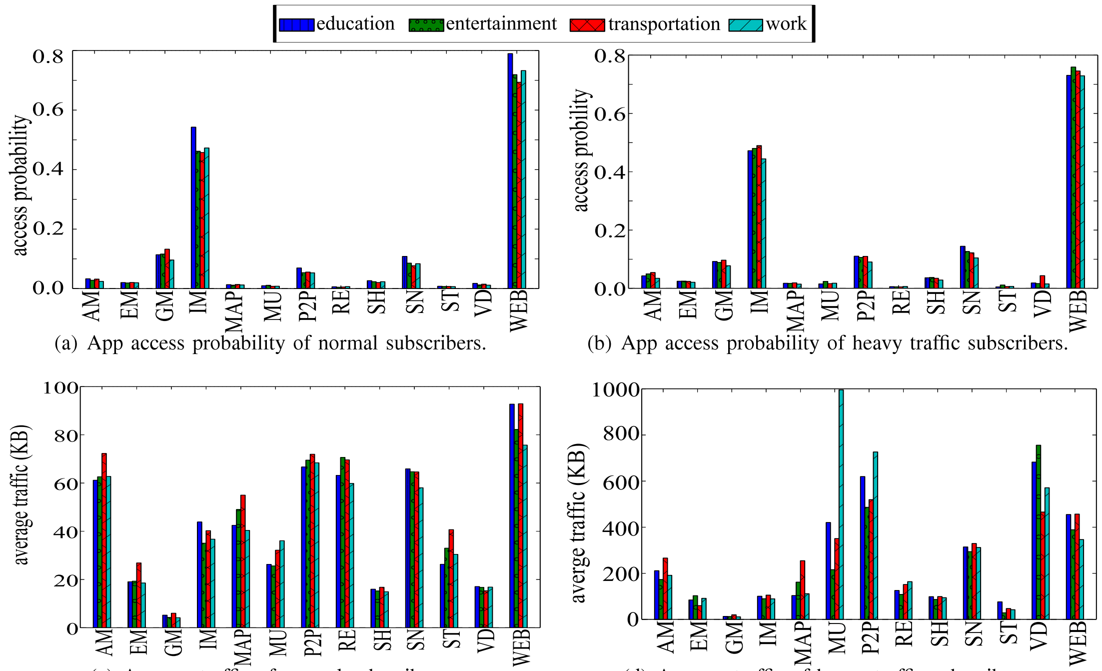
VI. App Usage Behavior of Subscriber Groups

In this section, we explore app usage behavior of two subscriber groups: **(a) heavy traffic subscribers**, defined as the top 20% of subscribers in terms of traffic volume. **(b) high mobility subscribers**, who are the top 20% of subscribers of the highest mobility. As both of these subscriber groups pose challenging requirements for network quality, one for network load and another for network connectivity, understanding of the app usage behavior of these subscribers groups could help cellular operators and app designers to provide better user experience.

A. Behavior of Heavy Traffic Subscribers

Heavy traffic subscribers are defined as the top 20% of subscribers according to their daily average traffic. As our data set contains approximately 8 million subscribers, there are about 1.6 million heavy traffic subscribers in this measurement.

First, we investigate the app interest of heavy traffic subscribers. Figure 7 shows the number of unique users and normalized traffic of heavy traffic subscribers on each app category. We notice that *web browsing*, *p2p* and *instant messager* are still the most popular apps among heavy traffic subscribers, *e.g.*, together they account for more than 80% of the total traffic.



(a) App access probability of normal subscribers.

(b) App access probability of heavy traffic subscribers.

(c) Average traffic of normal subscribers.

(d) Average traffic of heavy traffic subscribers.

Fig. 12. Impact of location on app usage of normal/heavy traffic subscribers.

Then, we also compare the mobility of heavy traffic subscribers with normal subscribers. Here, normal subscribers are defined as the remaining 80% of total subscribers excluding heavy traffic subscribers. Figure 8 demonstrates the mobility of these two groups of subscribers. According to this figure, the mobility of heavy traffic subscribers outstrips the mobility of normal subscribers. For example, 11% of heavy traffic subscribers move across more than 14 cells in a day, while less than 5% of normal subscribers have much higher mobility. A similar pattern can be observed in terms of radius of gyration. This finding corresponds to our previous observation in section IV, which indicates subscribers of higher mobility tend to generate more traffic.

Our previous study in section IV-B indicates that the impact of mobility on each app category varies, thus a follow-up question is whether mobility has a similar effect on heavy traffic subscribers. To answer this question, we compare the impact of mobility on the average traffic of normal subscribers and heavy traffic subscribers in Figure 9. Interestingly, the average traffic of normal subscribers increases dramatically with the number of visited cells, but in terms of radius of gyration, the average traffic first raises and then drops significantly. This implies that normal subscribers only tend to generate more traffic when they are travelling within a relatively small geospatial region, *e.g.,* commuting via public transport in downtown. Meanwhile, the patterns of heavy traffic subscribers are quite different: the average traffic slightly changes as the *number of visited cells* grows, but there exists a surprising increasing trend when the *radius of gyration* is larger than a certain threshold (13 kilometres in our dataset). Such difference suggests that heavy traffic subscribers always tend to generate more traffic, even when they are traveling over a long distance.

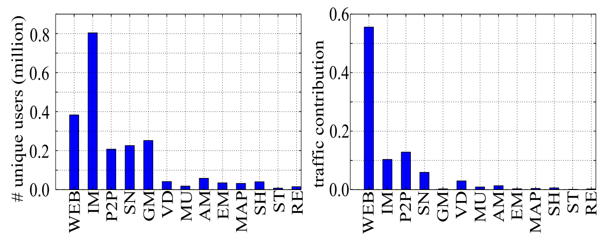
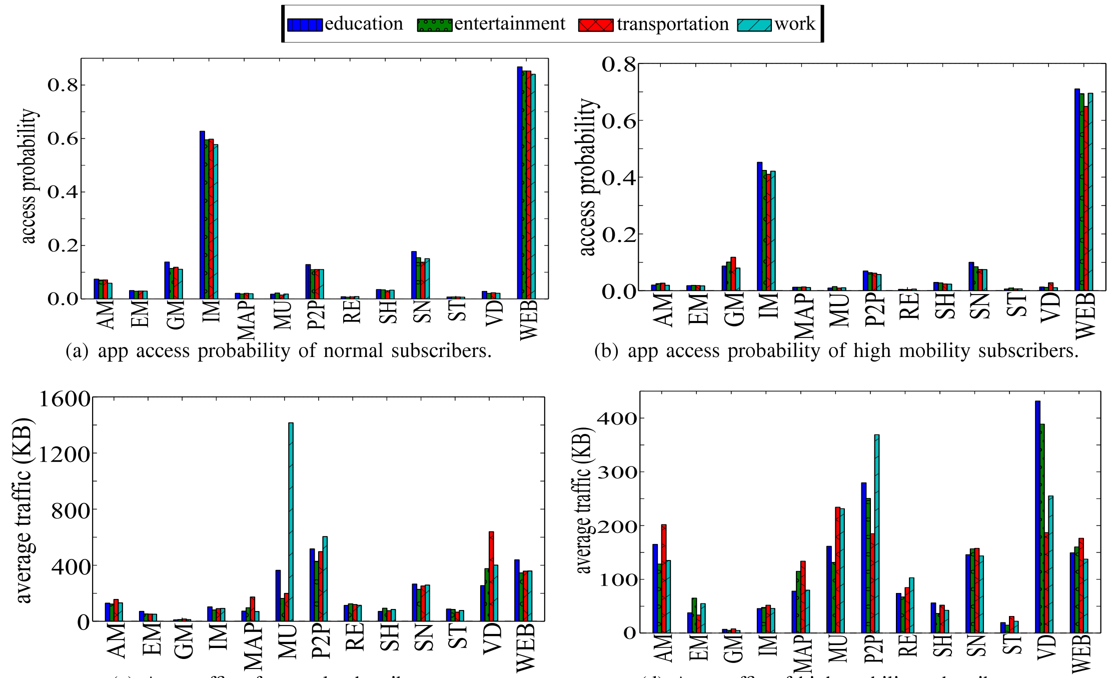


Fig. 13. App interest of high mobility subscribers

Figure 10(a) and Figure 11(a) show the mobility effect on the app access probability of normal subscribers and heavy traffic subscribers. We notice that, the access probability of *web browsing* and *instant messager* of heavy traffic subscribers remain almost 1, and *social network* and *game* are also relatively high.

According to Figure 10(b) and Figure 11(b), we find the traffic composition of normal traffic subscribers are relatively stable, only an increasing trend of *web browsing* in term of number of visited cells is observed. However, when referring to heavy traffic subscribers, the traffic of *video* and *music* traffic fluctuate dramatically as the radius of gyration grows, which may be due to the fact that these apps place a high requirement on network quality and thus more sensitive as mobility increases.

Next, we examine the impact of location on the app usage of heavy traffic subscribers. Figure 12(a) and 12(b) presents the app access probability of each app category in different locations. It is obvious that location merely affects app access probability for both normal and heavy traffic subscribers. However, the average traffic of each app category in different location in Figure 12(c) and 12(d) demonstrates the type of location significantly affects the app traffic of heavy traffic sub-



(a) app access probability of normal subscribers. (b) app access probability of high mobility subscribers.

(c) Avg traffic of normal subscribers. (d) Avg traffic of high mobility subscribers.

Fig. 14. Impact of location on app usage of normal/high mobility subscribers.

scribers, while barely having an effect on normal subscribers. For example, for heavy traffic subscribers, *music* and *P2P* only significantly dominates in *work* and *education* areas.

B. Behavior of High Mobile Subscribers

Finally, we examine the app usage behavior of high mobility subscribers. After ranking all subscribers according to their daily mobility, we define the top 20% subscribers of the highest mobility as high mobility subscribers and the remaining 80% of subscribers as normal subscribers.

As mobility affects network connectivity and quality, knowing what kind of apps are more frequently used by high mobility subscribers could help OS vendors and app designers to prepare for network variations caused by high mobility. Therefore, we show the app interest of high mobility subscribers in Figure 13. We notice that, except for a slight growth in *social network*, the interest of high mobile subscribers barely differs from other users. The fact that people tend to stay connected to others when they are out of their comfort zone, *e.g.*, home, sheds light on this trend [7]. In terms of traffic volume, *web browsing* dominates. This might be explained by the fact that *web browsing* does not require a long connection or intensive data transmission, thus it is more adaptable to the high mobility scenario.

Besides, we also wonder how location affects app usage for high mobility subscribers. We find that, similarly, location does not affect app access probability much, but has a significant effect on app usage extent, *i.e.*, app traffic. According to Figure 14, location has a very different effect on normal and high mobility subscribers. For example, in *education* areas, *video* contributes a large proportion of traffic for high mobile subscribers, but a relatively small proportion for normal subscribers.

VII. Implications

In previous sections, we analysis mobile app usage from mobility, spatial and subscriber groups perspectives. We believe our observations could have many implications. In this section, we briefly discuss some implications from cellular network operator, content provider and OS/App vendor perspectives.

A. Cellular Network Operator

As Mobility is one of the most import characteristics of cellular networks, we carefully investigate how mobility affects subscribers' traffic and app usage. These findings could help cellular operators to optimize their network performance. For example, we notice that the mobility level of heavy traffic subscribers is significantly higher than normal subscribers. As heavy traffic subscribers contribute most of the mobile traffic, cellular network operators could benefit from designing more efficient handover policy [15] for these subscribers.

Also, we notice there are some obvious spatial patterns for app usage. For example, *music* and *P2P* are well used in *work* areas. As such data-intense apps could pose higher network quality requirements, *e.g.,* higher bandwidth, cellular network operation could improve their service by optimizing resource allocation in such locations at the fine-grained level.

B. content provider

Content providers can take advantage of spatial patterns of app usage to optimize their service: In our analysis on spatial patterns of app usage, analysis results indicates that a strong spatial pattern exists for some apps, *e.g., music* dominates the traffic in *work* areas, and *video* is more frequently used in *entertainment* areas. These observations suggest a content provider could consider placing their content distribution network near these locations.

C. OS/App vendor

In our analysis on mobility patterns, we make an interesting observation that some apps are more frequently used and contribute more traffic when subscribers roam across more cells. For example, the access probability of *game* and *social network* shows significant growth as the number of visited cells increases. Since frequently roaming between multiple cells could cause dramatical network quality variations [12], the vendors of these apps should consider adopting technologies,such as caching and pre-fetching, to handle network quality variations.

VIII. Related Work

There exists a plethora of works studying cellular traffic and app usage from different perspectives. The authors in [3], [4] provide detailed analysis on app usage patterns and energy consumption based on detailed logs collected by pre-installed app on smartphone, but are limited by the scale of the data sets. [7] is a pioneering work which uses large-scale network trace logs to study the app interest and its geospatial characteristics. This work makes a very coarse partitioning of the cells into *home*, *work* and *hotspot*. Shafiq and *et al.* [14] focus on the geospatial pattern of app usage, in which they only consider locations as *downtown*, *university* and *suburb*. Compared with these works, our work takes a step further to investigate the location impact at a fine-grained level, we carefully investigate the impact of locations of different functions. Besides, [2], [8], [16]–[18] provide a broad view of traffic dynamics by analysing large-scale network data traces collected from 3G network. Instead of focusing on detailed app usage pattern, they put more emphasis on traffic dynamics from perspective of the cellular network provider. For us, we focus more on detailed traffic dynamics of different apps.

Another group of works focus on the optimization of cellular networks and mobile devices. [19] leverages machine learning algorithms to optimize power consumption, [20] proposes a method to maximize the utilization of bandwidth to reduce the battery drain caused by background communication, while [21] invents a mechanism to save power consumption of Base stations in 3G networks. Zang *et al.* proposed an efficient paging scheme to locate mobile devices quickly via mining large-scale call detail records in [22], Xu *et al.* developed a novel algorithm to accurately assigns IP performance measurements in [23], and the authors in [6], [24] inverted tools to analyse mobile app usage. We believe our work can provide some implications for further optimization, and complements these works.

IX. Conclusion

In this work, we investigated app usage patterns of mobile apps, mainly focusing on understanding how subscribers' mobility, geospatial properties and interest affect the usage patterns of mobile apps.

According to our measurements, we notice that, not only the subscriber's mobility pattern significantly diversifies, but also the impact of mobility on each mobile apps obviously varies. Our analysis on geospatial properties of mobile apps also suggests that location plays an important part to the extent to which a mobile app can be used. Besides, we identify two different groups of subscribers, *i.e.*, heavy traffic subscribers and high mobility subscribers, and investigate their preference and usage behavior of mobile apps respectively. We believe our study is at more fine-grained level and complements the knowledge gap in previous studies.

Acknowledgment

The research was supported in part by grants from 973 project 2013CB329006, RGC under the contracts CERG MHKUST609/13 and 622613, China NSFC under Grant 61502114, 61373092, 61272449, 61572339 and 61202029.

References

[1] C. V. Networking, "Cisco visual networking index: Global mobile data traffic forecast update, 20132018," *White Paper, February*, 2014.

[2] Q. Xu and et al., "Identifying diverse usage behaviors of smartphone apps," in *Proc. IMC*, 2011.

[3] H. Falaki and *et al*, "Diversity in smartphone usage," in *Proc. MobiSys*, 2010.

[4] H. Falaki and *et al.*, "A first look at traffic on smartphones," in *Proc. IMC*, 2010.

[5] G. Maier, F. Schneider, and A. Feldmann, "A first look at mobile hand-held device traffic," in *Proc. Passive and Active Measurement*, 2010.

[6] J. Huang and *et al.*, "Anatomizing application performance differences on smartphones," in *Proc. MobiSys*, 2010.

[7] I. Trestian and *et al.*, "Measuring serendipity: connecting people, locations and interests in a mobile 3g network," in *Proc. SIGCOMM*, 2009.

[8] U. Paul and *et al.*, "Understanding traffic dynamics in cellular data networks," in *Proc. INFOCOM*, 2011.

[9] S. Kumar and *et al.*, "Algorithms to accelerate multiple regular expressions matching for deep packet inspection," *ACM SIGCOMM*, 2006.

[10] W. C. Lee, *Mobile cellular telecommunications: analog and digital systems*. McGraw-Hill Professional, 1995.

[11] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, 2008.

[12] F. P. Tso and *et al.*, "Mobility: a double-edged sword for hspa networks," in *Proc. MobiHoc*, 2010.

[13] J. Lee Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician*, vol. 42, no. 1, 1988.

[14] M. Z. Shafiq and *et al.*, "Characterizing geospatial dynamics of application usage in a 3g cellular data network," in *INFOCOM*, 2012.

[15] 3GPP, "Self-configuring and self-optimizing network use cases and solutions," in *3GPP TR 36.902, Rel. 11*, Feb. 2012.

[16] Y. Jin and et al., "Characterizing data usage patterns in a large cellular network," in *Proc. CellNet*, 2012.

[17] M. Wittie and *et al.*, "Mist: Cellular data network measurement for mobile applications," in *BROADNETS*, 2007.

[18] R. Keralapura and *et al.*, "Profiling users in a 3g network using hourglass co-clustering," in *Proc. MobiCom*, 2010.

[19] P. K. Athivarapu and *et al.*, "Radiojockey: mining program execution to optimize cellular radio usage," in *Proc. MobiCom*, 2012.

[20] A. Chakraborty and *et al.*, "Coordinating cellular background transfers using loadsense," in *Proc. MobiCom*, 2013.

[21] C. Peng, S.-B. Lee, S. Lu, H. Luo, and H. Li, "Traffic-driven power saving in operational 3g cellular networks," in *Proc. MobiCom*, 2011.

[22] H. Zang and J. C. Bolot, "Mining call and mobility data to improve paging efficiency in cellular networks," in *Proc. MobiCom*, 2007.

[23] Q. Xu and *et al.*, "Acculoc: Practical localization of performance measurements in 3g networks," in *Proc. MobiSys*, 2011.

[24] F. Qian and *et al.*, "Profiling resource usage for mobile applications: A cross-layer approach," in *Proc. MobiSys*, 2011.